生命科学名著

J.D. 沃森 T.A. 贝克
〔美〕S.P. 贝尔 A. 甘恩 编著
M. 莱文 R.M. 洛斯克
杨焕明 等 译

科学出版社 生物分社
编辑推荐
Editor's Choice

（第六版）

基因的分子生物学

MOLECULAR BIOLOGY OF THE GENE

(Sixth Edition)

内含光盘

科学出版社
www.sciencep.com

PEARSON Education

《基因的分子生物学（第六版）》（含光盘）

Molecular Biology of the Gene

本书由DNA双螺旋的发现者之一——James D. Watson及其他几位著名学者在第五版的基础上修订完成。本书除了反映分子生物学领域的最新进展之外，还涉及其他诸多方面的内容。此次修订仍保留了上一版中的许多定义和特点，而且新增了调控RNA和基因组分析以及系统生物学的内容。全书共分五篇：化学和遗传学、基因组的维持、基因组的表达、调控以及方法。

本书具有权威性，内容新颖、详尽，堪称此领域的经典之作。为广大的生物爱好者及研究人员提供了分子生物学的知识框架和实验途径，并强调了基因科学对于整个生物领域的重要意义。

生物分社
联系电话：010-64012501
http://www.lifescience.com.cn
e-mail:lifescience@mail.sciencep.com

ISBN 978-7-03-024756-8

销售分类建议：生物

定价：128.00 元(含光盘)

生命科学名著

基因的分子生物学

（第六版）

Molecular Biology of the Gene

(Sixth Edition)

J. D. 沃森　T. A. 贝克

〔美〕　S. P. 贝尔　A. 甘恩　　编著

M. 莱文　R. M. 洛斯克

杨焕明 等　译

科学出版社

北京

图字：01-2008-3370 号

内 容 简 介

本书由 DNA 双螺旋的发现者之一——James D. Watson 及其他几位著名学者在第五版的基础上修订完成。本书除了反映分子生物学领域的最新进展之外，还涉及其他诸多方面的内容。此次修订仍保留了上一版中的许多定义和特点，而且新增了调控 RNA 和基因组分析以及系统生物学的内容。全书共分五篇：化学和遗传学、基因组的维持、基因组的表达、调控以及方法。

本书具有权威性，内容新颖、详尽，堪称此领域的经典之作。为广大的生物爱好者及研究人员提供了分子生物学的知识框架和实验途径，并强调了基因科学对于整个生物领域的重要意义。

图书在版编目(CIP)数据

基因的分子生物学/(美)沃森(Watson, J. D.)等编著;杨焕明等译. —北京:科学出版社,2009

(生命科学名著)

书名原文:Molecular Biology of the Gene

ISBN 978-7-03-024756-8

Ⅰ. 基… Ⅱ. ①沃…②杨… Ⅲ. 基因-分子生物学 Ⅳ. Q78

中国版本图书馆 CIP 数据核字(2009)第 096158 号

责任编辑:夏 梁/责任校对:刘小梅

责任印制:钱玉芬/封面设计:陈 敬

科学出版社出版

北京东黄城根北街 16 号

邮政编码：100717

http://www.sciencep.com

骏杰印刷厂印刷

科学出版社发行 各地新华书店经销

*

2009 年 7 月第 一 版 开本:787×1092 1/16

2009 年 7 月第一次印刷 印张:53 1/2

印数:1—3 000 字数:1 221 000

定价:128.00 元(含光盘)

(如有印装质量问题,我社负责调换〈环伟〉)

本书译校人员

（按姓氏汉语拼音排序）

冯小黎	郭需石	胡建飞	胡少晖
胡松年	李京湘	李　雷	刘　斌
刘国振	刘明旭	刘　韧	娄晓敏
时　亮	宋　光	宋其峰	宋述慧
苏夜阳	汪　浩	王彩平	王敦梅
王　霞	吴东颖	吴　琳	闫春霞
杨焕明	张　明	张清润	张　欣
赵　辉	朱金桂		

中 译 版 序

《基因的分子生物学》（第六版）的中文版能紧随其英文版迅速面世，使我非常高兴。

我们对基因的认识能以前所未见的速度更新，首先要归功于 DNA 测序技术的突破。调控 RNA 的新发现及其在细胞中操作的新技术也同样功不可没。

我的这本书，仍是由杨焕明教授领导他的团队翻译成中文的。他们作为分子生物学家的素质和水平，使我相信中国的同事和同学阅读本书中文版之所得，不会亚于阅读本书的英文版。

本书新版之日，适逢几千个人类个体的基因组正准备被测序之时。我衷心希望我的这本书能激励更多的年轻的中国科学家脱颖而出，参与并领导这一人类文明的伟大进步。

J. D. 沃森

2008 年 7 月 14 日于冷泉港

Foreword

I am delighted that the sixth edition of *The Molecular Biology of the Gene* is so soon appearing in a Chinese edition. Our knowledge about genes is moving ahead faster than ever before – thanks to the availability of powerful new technologies for reading the sequences of DNA molecules and expressed RNA molecules. Equally important has been the discovery of new classes of regulatory RNAs and methodologies for manipulating their presence within cells through the introduction of RNA molecules.

That my book's translation into Chinese is again done by a team of high quality molecular biologists led by Professor Huanming Yang reassures me that the sentences read by the Chinese university science students have identical meaning to those learned by students who read this book in English.

As we are about to start looking at the genomes of many thousands of human beings, I hope my book helps position many young Chinese scientists to help lead this major step forward for human civilization.

James D. Watson

14 July 2008
Cold Spring Harbor, New York

译　者　序

《基因的分子生物学》（第五版）自2003年问世以来，好评如潮。2008年发行的第六版，又“百尺竿头，更进一步”。

与前五版一样，第六版保留了本书的风格和结构。但与第五版相比，第六版不仅对几乎所有章节进行了精心修订（有几章的修改幅度还很大），而且新增了第18章调控RNA和第20章基因组分析与系统生物学，以反映这5年来生命科学领域的重要进展。

在较为仔细地阅读了本版新增的这两章之后，公平地说，这新的两章毕竟是初出茅庐，略显粗糙：测序技术的描述在刚出版不久的今天已有过时之嫌。这一方面是由于任何一本书的再版速度也不可能赶上测序技术这几年的“革命性”的飞速进展；另一方面，该章的执笔者也似乎未必是测序圈内的“弄潮儿”。而有关“系统生物学”那一章的不尽如人意，责备撰写者也许是不公平的。正如作者在前言中特意说明的，这一时尚的新学科的名称也“尚未明确定义（rather ill-defined）”。名不正，言就难以很顺。相比而言，调控RNA的部分可能是写得最好的，也许是我恰恰对这一领域了解最少。

这一点也不奇怪。这正说明了本书为什么近45年来一直经久不衰，备受青睐。经过6次洗礼的本书原有章节，已经千锤百炼，堪称炉火纯青，绝非一日之功。我对新增两章的评论，也是出于“爱之深，责之切”，一点也不影响我对这两章以至于全书进一步提高的信心。我已为本版所体现的作者精益求精的态度所折服。

我热忱地向读者推荐本书的第六版。这不仅是由于第六版的出版周期比前一版缩短不小，而且是由于本版的翻译质量有了一定的提高。在审阅本版时，我深为前一版中的一些错误和不妥之处感到惭愧，也为本版肯定同样存在的瑕疵而提前向读者致歉。

我要特别感谢Watson博士。感谢他在百忙之中为本书中文第六版写的序言。我至今不能忘记，那正是他一生中也许唯一的多事之秋。文如其人。让我们一起来细细品味这一短文，来好好理解这位科学巨匠在耄耋之年对这一领域的期待，特别是对中国年轻一代的期望：“我衷心希望我的这本书能激励更多的年轻的中国科学家脱颖而出，参与并引领这一人类文明的伟大进步。”

杨焕明

2009年5月于深圳

前　言

自四年前《基因的分子生物学》第五版问世以来，分子生物学发生了三个重大的变化：第一是基因组测序成本的降低及其深远的影响；第二是 RNA 作为调控分子的认识及其广泛的认可；第三是被好多人奉为基因组时代之后分子生物学未来的系统生物学。本书第六版充分反映了所有这三方面进展的影响。

在过去的四年中，我们已经完成了很多生物的基因组序列，其中还包括本书一位作者的序列。现在，一个基因组的测序成本只有 2003 年的五百分之一。那一年，我们只有模式生物——酵母、线虫、拟南芥、果蝇、小鼠、人类的基因组序列，或许还有其他两三种动物。现在，我们有了更多的生命之树的代表物种的基因组序列：黑猩猩、大鼠和蜜蜂；猫和狗；好多种昆虫；水稻，海葵，负鼠（opossum）；和许多其他基因组。我们在同一类生物中也有了更多的代表性基因组：例如，到目前为止已有 13 个种的果蝇基因组序列被测定。更为重要的是，我们已经进入了个体基因组的时代。同已提到的 James Watson 一起，Craig Venter 的基因组也已经完成测序。

序列数量的爆炸不仅使测序技术的影响随处可见，也使测序成为生物实验室的常规技术。本书这一新版本中的好多进化和机理的讨论，其论点也是基于或近缘或远缘的物种的基因组比较数据。

正是基因组序列带来了第二个重大的进展——对 RNA 全部功能的认识。替换剪接的重要影响和调控 RNA 的广泛应用的内容，代表了这一版本的最大修订。

系统生物学仍然是一个定义不明（ill-defined）的术语，不同的人有不同的理解。一些人认为这只不过是高通量的方法学，另一些人则看成是生物系统的数学模型。在本版中，我们讨论了这一新兴领域与本书内容最明显最直接相关的一个方面：基因调控网络的阐述和模型建立。

本书的结构

尽管有很多修改，但本书的很多核心价值和结构框架并没有改变。因而，新版的结构仍与上一版本一样，分为五个部分：

第一篇　安排了全书所有内容的基础知识，叙述了遗传和分子生物学的历史，解释了决定大分子结构和功能的化学原理。

第二篇　涵盖了遗传物质的组织和维持。从介绍 DNA 和 RNA 结构入手，到组装到染色体中的基因组组成，解释了 DNA 复制，重组和修复。

第三篇　阐述了通过转录、剪辑和翻译过程的基因组表达。

第四篇　讨论了基因组中基因表达的调控。包括转录调控的机制，调控 RNA 在基因表达的作用，进化中调控机制的演变。最后，我们安排了基因表达的全基因组范围的

研究，系统生物学和合成生物学等新兴领域中基因调控网络的模型建立和分析。

第五篇 由两章组成：一章是分子生物学、基因组学、蛋白质组学和生物信息学的技术，另一章是已经给我们揭示了许多分子生物学基本原理的模式生物的研究。

新版不仅对每一章都做了修改，而且还增加了好多新的内容。最显著的是新增了两个新的章节和一系列新的框图。

新章节和框图

18 章：调控 RNA　自第五版以来，最显著的进展是发现数目众多的 RNA 都是基因的调控因子。在该章中，我们差不多讨论了所有相关的发现：从细菌的核糖开关（riboswiches）和小 RNA（small RNA），到真核生物的 RNA 干扰（RNA interference）和 microRNA，再到 X 染色体失活中 RNA 调控子的作用等等。除了详细讨论这些调节因子的工作机制外，我们还讨论了它们是如何被发现的，又是如何给我们提供了基因表达人工操作的新途径。

20 章：基因组分析和系统生物学　通过许多动物的基因组测序发现它们的基因组序列高度相似。这个发现强调了一个基因组中决定表型差异的绝大部分基因在调控上的区别。现在我们有了研究全基因组的基因表达方式的技术，因而在此讨论这些技术的工作原理和基于这些技术的发现。在本章的第二部分，我们还讨论了关于基因调控网络的代表性“线路图”（wiring diagrams），突出地说明了这些途径中的信息流动和“决策”方式。

框图　第六版中增添了一系列新的框图，分别表示：

- **关键实验**　描述了在各个领域中的历史性的和新近的关键实验。在不打断正文的文字表述的同时，给学生介绍科学问题是如何通过实验来解决的。
- **技术**　描述了与该章节相关的技术，与本书最后两章描述的常见技术互补。
- **医学联系**　强调了基础分子生物学和多种人类疾病之间的联系，重点强调了基础研究如何推进现代医学的发展。
- **最新概念**　在更高层次上探究所选择的一些新概念，加深学生对前沿思想的认识。

冷泉港实验室照片　正如前面所有版本一样，本版的每一篇都以一组照片开始，其中一些是本版新增的。这些照片都选自冷泉港实验室档案馆，并且都是在冷泉港拍摄的，大部分是自 1933 年开始几乎每年夏天都在那里举行的专题讨论会的照片。每一张照片的说明都标出了这些人物及拍摄的时间。更多的历史性的镜头可以浏览冷泉港实验室档案馆的网站（http://archives.cshl.edu/）。

致谢

本版中的部分内容采用了作者之一（Richard Losick）在哈佛大学教授的分子生物学基本教程。作者要感谢过去几年中为该教程做出贡献的 Steve Harrison 和 Jim

Wang，他们的影响体现在第六章及其他一些章节的内容之中。我们还要特别感谢Craig Hunter撰写了第21章中有关线虫的部分，以及Rob Martienssen撰写的有关植物拟南芥的部分。

本书的部分手稿曾呈送很多同事审阅，他们宝贵的意见对于保证本书的内容及插图能够反映潮流和趋势功不可没。我们特别感谢Katsura Asano，Jamie Cate，Amy Caudy，Richard Ebright，Mike Eisen，Chris Fromme，Brenton Graveley，Ann Hochschild，JimHu，Richard Jorgensen，David Jeruzalmi，Leemor Joshua-Tor，Sandy Johnson，Adrian Krainer，Karoline Luger，Julian Lewis，Sue Lovett，Rob Martienssen，Bill McGinnis，Matt Michael，Lily Mirels，Nipam Patel，Mark Ptashne，Danny Reinberg，以及Bruce Stillman.

我们还要感谢所有本书插图的提供者和创意、制作者：Sean Carroll，Seth Darst，Richard Ebright，Brenton Graveley，Ann Hochschild，Julian Lewis，Bill McGinnis，Phoebe Rice，Dan Rockhsar，Nori Satoh，Matt Scott，Peter Sorger，Tom Steitz，Andrzej Stasiak，Dan Voytas和Steve West.

我们还要特别感谢Leemor Joshua-Tor以她的高超技术和超常耐心绘制了书中所有出色的结构插图。我们同时也要感谢本书的软件提供者①：Per Kraulis，Robert Esnouf，Ethan Merrit，Barry Honig和Warren Delano。我们还得到了蛋白质数据库（Protein Data Bank，www. rcsb. org/pdb/）的协助，在插图的说明里我们特别标注了每一个结构的解析者。

我们的美术工作仍是由Dragonfly传媒公司的一群天才艺术家在Craig Durant的领导及Helen Wortham的帮助下完成的。Denise Weiss和Ed Atkeson制作了本书的封面。我们感谢Clare Clark和CSHL档案馆所提供的部分草图和为寻找到这些图所付出的努力。

我们感谢Benjamin Cummings的Susan Winslow，她为本版的问世做了很多工作，更多的是解决那些棘手的问题。我们感谢Gary Carlson在后期接手这一工作，他使所有的工作得以顺利继续。我们感谢Jan Argentine监管本书的编写和出版的全过程，他帮助我们解决所出现的问题，重新组织工作和时间，满足我们所有的需求，并保证了我们的神圣时间表的执行。我们感谢Kaaren Janssen对本书的编辑，重写我们凌乱的初稿，精力充沛，孜孜不倦。Inez Sialiano对草稿的整理和校正，以及Carol Brown的所有支持。在出版过程中，Kathleen Bubbeo和Susan Schaefer表现出来的极大的耐心，

① MolScript软件得到Per Kraulis的允许使用（Kraulis P. J. 1991. MOLSCRIPT：A program to produce both detailed and schematic plots of protein structures. Journal of Applied Crystallography 24：946-950）。BobScript软件经Robert Esnouf允许使用（Esnouf R. M. 1997. Journal of Molecular Graphics 15：132-134）。此外，Raster3D软件得到了Ethan Merritt的允许使用（Merritt E. A. 及Bacon D. J. 1997. Raster3D：Photorealistic Molecular Graphics. Methods in Enzymology 277：505-524），Barry Honig允许我们使用GRASP软件（Nicolls A.，Sharp K. A. 及Honig B. 1991. Protein folding and association：Insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins 11：281-296）。PyMOL软件得到Warren DeLano的允许使用（DeLano W. L. 2002. The PyMOL Molecular Graphics System. DeLano Scientific，Palo Alto，California）。

只有这样的耐心才使我们能够对初稿进行不断的修改，直到我们以为已经完成的最后时刻。我们也十分感谢 Denise Weiss，她不仅负责整个过程，还亲自作了很多版面设计的工作而使本书锦上添花。John Inglis 一直监督本书的进程，并在需要的时候提供很多有用的建议。

最后，我们再一次感谢我们的家人和朋友，感谢他们给予的有力支持和对我们的有求必应。

JAMES D. WATSON
TANIA A. BAKER
STEPHEN P. BELL
ALEXANDER GANN
MICHAEL LEVINE
RICHARD LOSICK

（宋述慧　王　霞　译　刘　韧　杨焕明　校）

作 者 简 介

James D. WATSON 1968～1993 年任冷泉港实验室主任，1994～2003 年任董事长（President），现为教务长（Chancellor）。他在芝加哥大学接受大学教育，1950 年获印第安纳大学博士学位。1950～1953 年间在丹麦哥本哈根及英国剑桥从事博士后研究。在剑桥期间，他开始了与 Crick 的合作，由于这一合作才有了 1953 年阐明的 DNA 双螺旋结构（也正是因为这个发现，Watson，Francis Crick 和 Maurice Wilkins 在 1962 年获诺贝尔奖）。1953 年下半年，Watson 在加利福尼亚理工学院工作，1955～1976 年在哈佛大学从事教学以及 RNA 和蛋白质合成的研究工作。1989～1992 年担任美国国立卫生研究院（NIH）的国家基因组研究所的首任所长，2007 年测定了他自己的基因组。Watson 博士是《基因的分子生物学》一书第一版、第二版、第三版的作者，第四版和第五版的合著者。前五版分别于 1965 年、1970 年、1976 年、1978 年和 2003 年出版。Watson 还参与了其他两本教科书《细胞的分子生物学》和《DNA 重组》的写作。

Tania A. BAKER 现为麻省理工学院（MIT）的 Whitehead 教授及 Howard Hughes Medical Institute 的研究员。她在威斯康星大学（Madison）获生化学士学位，1988 年在斯坦福大学获生化博士学位。研究生期间在 Arthur Kornberg 教授的实验室工作，研究方向为 DNA 复制的起始机制。在 NIH 从事博士后研究期间，她在 Kiyoshi Mizuuchi 博士的实验室进行 DNA 转座的调控机制研究。她目前的研究是探讨遗传重组的机制和调控、酶催化的蛋白质去折叠和依赖 ATP 的蛋白质降解。Baker 教授曾荣获美国微生物学会颁发的 2001 年 Eli Lilly 科技奖及 MIT 理学院的 2000 年大学生教学奖。2004 年当选为美国艺术科学院（American Academy of Arts and Sciences）院士，2007 年当选为美国国家科学院（National Academy of Sciences）院士。与 Arthur Kornberg 合著《DNA 复制》一书第二版。

Stephen P. BELL 现为麻省理工学院（MIT）的生物学教授及 Howard Hughes Medical Institute 的研究员。在西北大学获得生物化学系、分子生物学系、细胞生物学系和跨学科（Integrated Science Program）学士学位。1991 年获加利福利亚大学伯克利分校（UC Berkeley）的生物化学博士学位。研究生期间在 Robert Tjian 博士的实验室研究真核基因的转录。在冷泉港 Bruce Stillman 博士的实验室从事博士后研究，主要研究真核 DNA 复制的起始。目前专注研究真核染色体复制的调控机制。Bell 教授于 2001 年获 ASBMB-Schering Plough 科学成就奖，1998 年获 MIT 的大学生教学奖（Everett Moore Baker Memorial Award），2006 年获 MIT 理学院的教育奖。

Alexander GANN 现为冷泉港实验室出版社的主任编辑及冷泉港实验室 Watson 生物科学学院的在职教师。他在伦敦大学学院（UCL）获微生物学学士学位，1989 年获爱丁堡大学分子生物学博士学位。研究生期间在 Noreen Murray 实验室研究限制性内切酶的 DNA 识别。博士后期间在哈佛大学 Mark Ptashne 实验室研究转录调控，后在

UCL 的 Ludwig 肿瘤研究所 Jeremy Brockes 实验室研究蝾螈肢体再生。1996～1999 年任英国 Lancaster 大学讲师，随后来到冷泉港实验室。与 Mark Ptashne 合著《基因与信号》一书（2002 年）。

Michael LEVINE 现为 UC Berkeley 分子及细胞生物学教授，整合基因组中心的副主任。于 UC Berkeley 遗传学系获学士学位，1981 年获耶鲁大学分子生物物理及生物化学系的博士学位，其导师为 Alan Garen。1982～1984 年期间，在 Waler Gehring 及 Gerry Rubin 的实验室里从事博士后研究，主要研究果蝇发育的分子遗传学。Levine 教授的研究组目前的研究兴趣集中在果蝇及海鞘胚胎原肠胚形成有关的基因网络。他目前是 UC Berkeley 的 F. Williams 遗传学及发育学基金会主席。1996 年被授予美国国家科学院颁发的分子生物学 Monsanto 奖，1996 年和 1998 年分别当选为美国艺术科学院和国家科学院院士。

Richard M. LOSICK 现为 Maria Moors Cabot 生物学教授、哈佛学院教授及哈佛大学艺术及理学院的 Howard Hughes Medical Institute 教授。他在普林斯顿大学化学系获学士学位，在 MIT 获生物化学博士学位。刚读完研究生，他便成为哈佛教师学会的年轻教师（Junior Fellow），开始研究 RNA 聚合酶及细菌基因转录的调控。Losick 教授曾任哈佛大学细胞及发育生物学系和分子及细胞生物学系的主任，曾获 Camille & Henry Dreyfuss 教师-学者奖。Losick 教授是美国国家科学院和美国艺术科学院院士、美国科学进步促进会（AAAS）成员、美国微生物学院（American Academy of Microbiology）成员、美国哲学学会（American Philosophical Society）成员、Phi Beta Kappa 学会访问学者，于 2007 年获美国国家科学院 Selman A. Waksman 微生物学奖。

简　　目

内 容 详 注

第 1 篇　化学和遗传学

第 8 章 DNA 的复制 195

第 9 章 DNA 的突变和修复 260

第 10 章 分子水平上的同源重组 286

第 11 章 位点特异性重组和 DNA 转座 319

第 3 篇 基因组的表达

第 12 章 转录机制 379

第 13 章 RNA 剪接 418

第 14 章 翻译 460

第 4 篇 调　　控

第 16 章 原核生物的基因调控 545

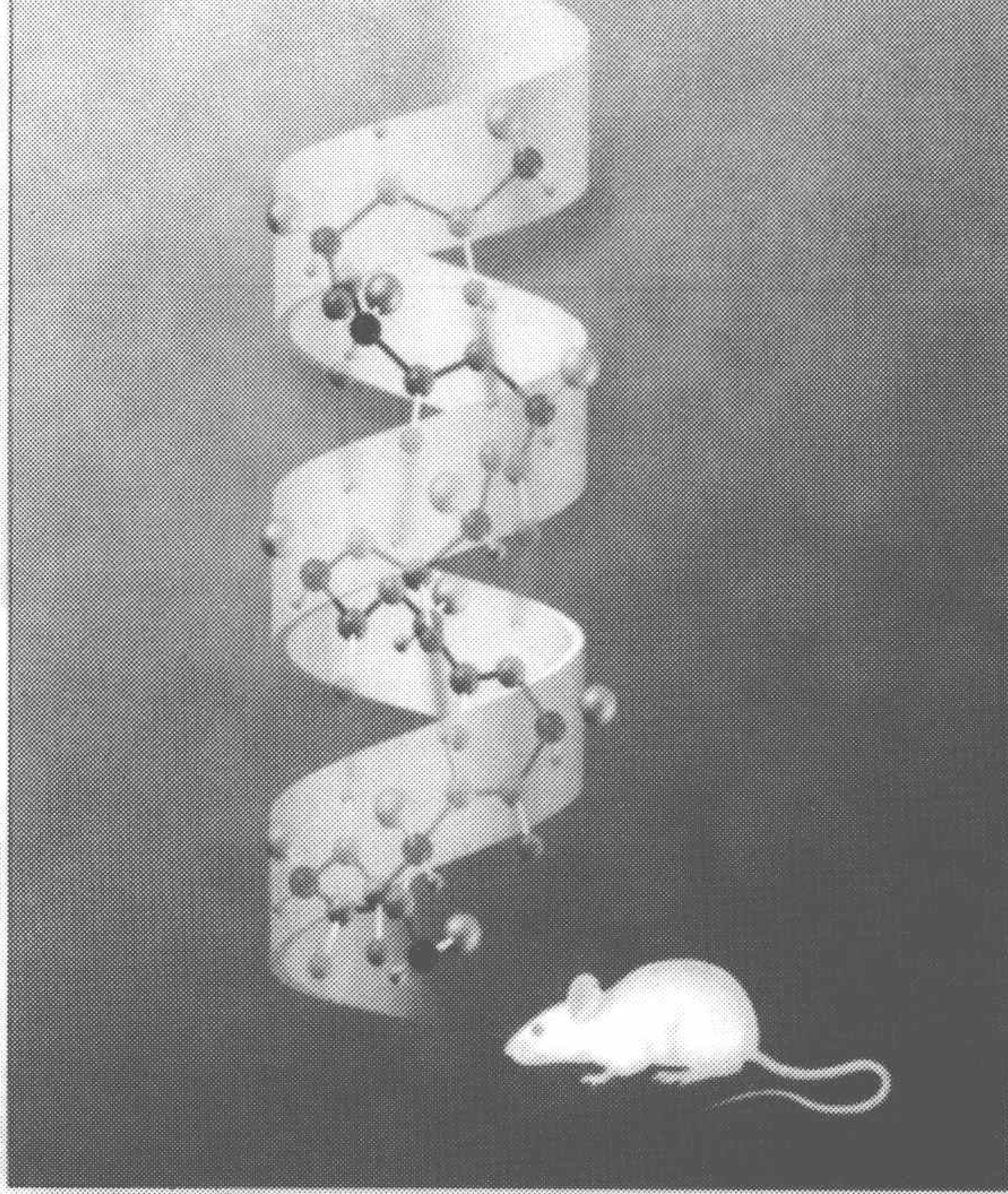
第 17 章 真核生物的基因调控 585

第 18 章 调控 RNA 631

第 20 章 基因组分析与系统生物学 705

第 5 篇 方 法

第 21 章 分子生物学技术 741



第1篇

化学和遗传学

本篇概要

- 第 1 章 孟德尔学派的世界观
- 第 2 章 核酸承载遗传信息
- 第 3 章 弱化学作用的重要性
- 第 4 章 高能键的重要性
- 第 5 章 弱键和强键决定大分子的结构

与本书其他章节不同，组成第 1 篇的 5 个章节所包含的内容与以前的版本相比基本上没有太大的变动。因为即使在进入基因组测序时代的当前，这一部分内容仍然同过去一样重要。具体来说，第 1 章和第 2 章主要对遗传学及其分子基础做一历史性的回顾，介绍了关键的思路和实验；第 3～5 章介绍了位于分子生物学核心地位的化学问题，我们将讨论构成大分子结构的基本化学原理，本书的主角正是这些大分子：DNA、RNA、蛋白质以及它们之间的相互作用。同以前的版本相比，主要的内容虽然没有太大的变化，但还是有一些新的编排，也增加了一些新的例子。

第 1 章着重介绍遗传学历史上的奠基性事件，我们将详细讨论从发现遗传学基本规律的著名的孟德尔豌豆实验，到 Garrod 的“一个基因一个酶”假说。第 2 章介绍了其后分子生物学的革命性事件，这一革命始于 Avery 的 DNA 是遗传物质的发现，继而 Watson 和 Crick 提出的 DNA 双螺旋模型和破译遗传密码以及“中心法则”(DNA 通过生成 RNA 而生成蛋白质)。这一章的结尾讨论了许多生物体的全基因组测序而产生的最新进展及测序对现代生物学的影响。

第 3～5 章所介绍的化学基础主要集中于化学键（强键和弱键）的性质以及它们在生物学中扮演的角色。

第 3 章的讨论从弱的化学作用开始，也就是氢键、范德华力（van der Waals）以及疏水作用。这些力是大分子之间，如蛋白质与蛋白质之间、蛋白质与 DNA 之间的相互作用的主要决定力。对大部分细胞活动和调控而言，弱键都是至关重要的。酶通过弱的相互作用与它们的底物结合；转录调控子结合到 DNA 的某些位置控制基因的开关也是通过弱相互作用起作用的。

单独的弱相互作用确实很弱，因此形成后往往可以很快地解离，这种可逆性对它们在生物学中的作用也很重要。在细胞内，分子间的结合必须是动态的（可逆的），不然整个系统就会“冻结”，同时，其中的一些相互作用至少在短时间内又必须是稳定的。为了满足这些看上去互相矛盾的要求，需要多种弱相互作用的协同使用。

强键将组成大分子的各个元件结合在一起。蛋白质是由特定顺序排列的、通过强键连接起来的氨基酸组成的，而 DNA 同样是由强键连接的核苷酸组成的（形成氨基酸和核苷酸的原子也是靠强键组合起来的）。这些键将在第 4 章介绍。

在第 5 章我们会看到强键和弱键如何共同决定了各个大分子独特的三维结构（以及

由此产生的特定功能)。正如弱键介导了大分子间的相互作用，一个蛋白质内不相邻的氨基酸之间也会起作用。由此，决定了氨基酸的初级顺序是如何折叠成三维形状的；同样地，DNA 分子的两条链也是靠弱键结合在一起的。

在第 5 章，我们还涉及了蛋白质功能的调控。一种方式是通过改变蛋白质的形状，也就是所谓的“变构”(allosteric) 调控。这样，一个蛋白质处于一种构象时可能会行使某种特定的酶学功能或者结合某种特定的靶分子；而处于另外的构象时则可能失去这些功能。这种形状的改变可以通过与其他蛋白质或小分子(如糖)的结合而引发；另外异构效应也能由共价修饰诱导，例如，蛋白质上结合一个或多个磷酸基团可以引发其形状改变。控制蛋白质的另外一种方式是控制它们与靶分子接触，通过这种方式某种蛋白质可以根据不同的信号而作用于不同的靶蛋白。

冷泉港实验室资料库中的照片



Vernon Ingram、Marchall Nirenberg 和 Matthias Staehelin，1963 年大分子的合成与结构学术会议。 Ingram 证明了基因控制蛋白质的序列，引起镰状细胞贫血症的突变是血红蛋白的一个氨基酸发生了变化(第 2 章)；Nirenberg 是破译遗传密码的主要完成者，他在体外用人工 RNA 模板引导了蛋白质合成(第 2 章和第 14 章)，因为这一成绩，他与其他人分享了 1968 年的诺贝尔生理学或医学奖；Staehelin 则从事小分子 RNA——tRNA的研究，tRNA 负责将遗传密码翻译成蛋白质的氨基酸序列(第 2 章和第 14 章)。



Raymond Appleyard、George Bowen 和 Martha Chase，1953 年的病毒学学术会议。这张照片中的 Appleyard 和 Bowen 都是噬菌体遗传学家。此外，照片中还有 Chase，她在 1952 年与 Alfred Hershey 合作，通过一个简单的实验使大多数人相信遗传物质是 DNA（第 2 章）。



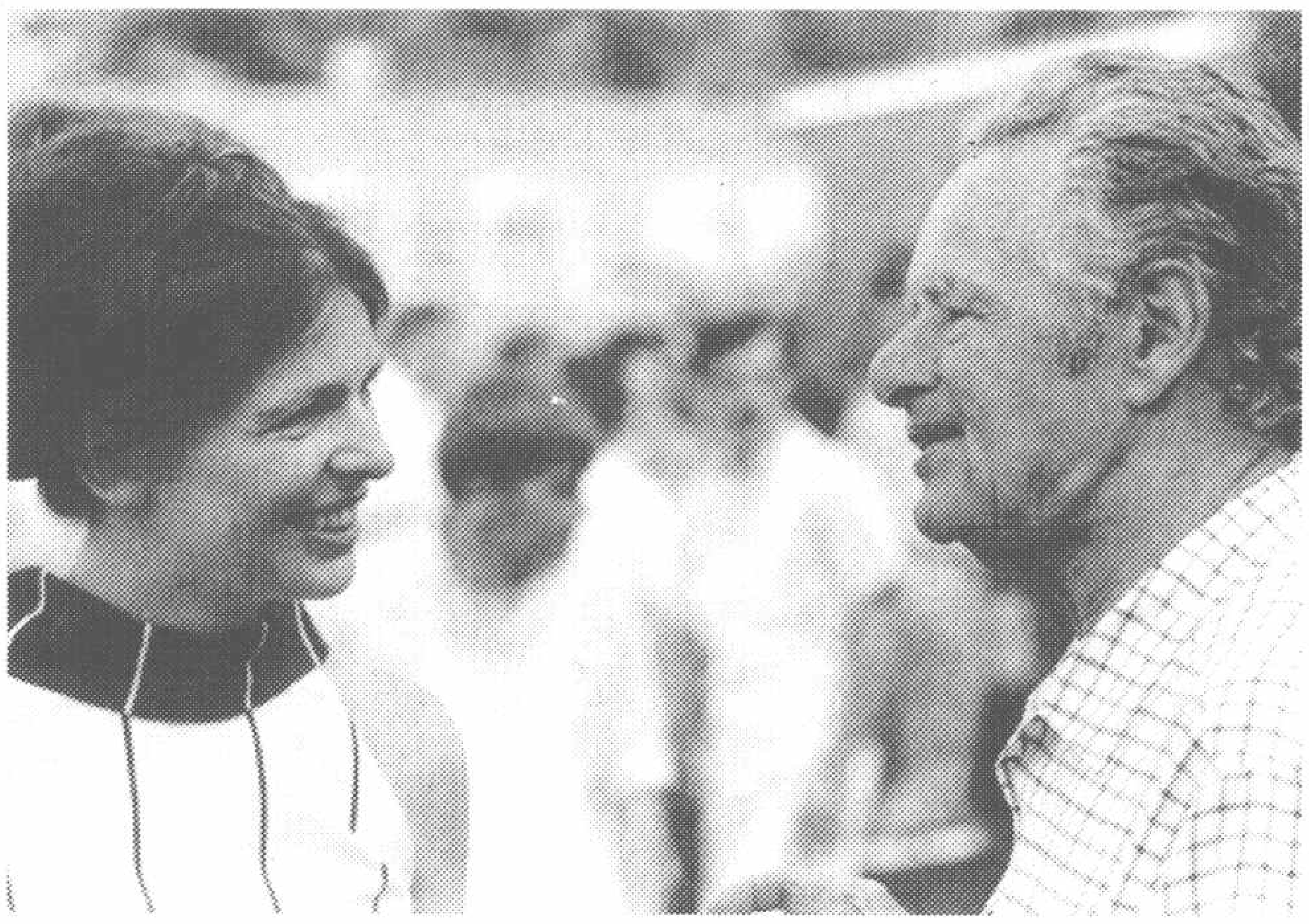
Melvin Calvin、Francis Crick、George Gamow 和 James Watson，1963 年的大分子合成与结构学术会议。因为在植物 CO_2 同化方面的工作，Calvin 获得了 1961 年的诺贝尔化学奖；Crick 和 Watson 因为提出了 DNA 的结构模型，分享了 1962 年的诺贝尔生理学或医学奖（第 2 章和第 6 章）；Gamow，一个被遗传密码的问题所吸引的物理学家（第 2 章和第 15 章），发起成立了一个由志同道合的科学家组成的非正式的组织——RNA 领带俱乐部（照片中他就系着一条由他本人设计的该俱乐部的标志领带）。



Calvin Bridges，1934 年的生长相关问题学术会议。Bridges（照片中读报者）是著名的摩尔根“果蝇小组”成员，他们开创性地发展了用果蝇（*Drosophila*）作为遗传模式生物的工作（第 1 章和第 22 章）；和他在一起的是 Buckholtz 博士。



Max Perutz，1971 年的三维水平上蛋白质的结构与功能学术会议。 Perutz 与 John Kendrew 分享了 1962 年的诺贝尔化学奖。经过长达 25 年的努力，他们终于用 X 射线结晶法（crystallography）首先分别解析了血红素和肌红蛋白的原子结构（第 5 章）。



Joan Steitz 和 Fritz Lipmann，1969 年的蛋白质合成机制学术会议。 Steitz 的研究集中在 RNA 分子的结构和功能，尤其是与 RNA 剪接有关的工作（第 13 章），她也是本书第四版的作者之一；Lipmann 证明了 ATP 中的高能磷酸基团是驱动许多生物反应的能量来源（第 4 章）。因为这一工作，他与 Hans Krebs 分享了 1953 年的诺贝尔生理学或医学奖。

第1章 孟德尔学派的世界观

本章概要

•

孟德尔的发现

•

遗传的染色体理论

•

基因连锁和交换

•

染色体定位

•

突变引起遗传变异

•

早期关于基因是什么以及如何发挥作用的推测

•

发现基因与蛋白质相互关系的初步尝试

人类在所有的生物中是独一无二的，这很容易理解。人类已经独自建立了复杂的语言系统，能够进行有意义的、复杂的思维和情感的交流，并且建立了伟大的文明进而改变了这个世界，这对其他的生命形式是不可思议的。所以，我们总是想当然地认为人类与其他物种相比，肯定有某种特殊性。这种观点在许多宗教形式中都有清晰的体现，通过这些宗教我们试图解释生命的起源以及我们存在的理由，进而尝试创造一种可操作的、用来规范我们生活的条律。在一个多世纪以前，人们很自然地认为，就像每个人的生命有特定的起始和终结时间一样，整个人类及其他所有生命形式应该也是在一个特定时刻被创造出来的。

140年以前，当达尔文（Charles Darwin）和华莱士（Alfred R. Wallace）基于适者生存的观点提出了进化理论时，这种"创世论"观点才第一次受到了真正的质疑。他们认为，各种形式的生命不是恒定不变的，而是持续地产生一些略有不同的动植物，其中的一些因为适应环境而存活下来，并且产生更多的后代。在进化论提出的时代，他们还不知道这些持续发生的变异的起因，但是他们确实正确地意识到，如果这些变化要想成为进化的基础，由此所获得的新特征一定要在后代中保持下去。

最初，达尔文受到了猛烈的抨击，主要是来自那些不愿意相信人类与丑陋的猿类可能有共同祖先的人们，哪怕这些祖先是生活在一千万年以前。也有一些反对者是生物学家，他们认为达尔文的证据不能令人信服。其中一位是著名的自然学家叫 Jean L. Agassiz,他当时在哈佛大学，花了多年的时间写文章反对达尔文和他的伙伴赫胥黎(Thomas H. Huxley)，赫胥黎是当时最有影响的进化论的推行者。到19世纪末期，科学上的争论已经基本结束了。如今动植物的地理分布以及在化石记录中的选择性地出现(selective occurrence) 只能有一种解释，那就是连续进化的生物来自共同的祖先。今

天，进化论已经被普遍接受，当然除了极少数教义派人士（fundamentalist），他们拒绝的理由不是基于理智，而是来自宗教的教条。

根据达尔文的理论，我们可以立即推论，40 亿年以前存在于地球上的生命是一种简单的形式，也许就像我们今天所知的最简单的生命形式——细菌，存在这样小的细菌也告诉我们生命状态的实质是在很小的生物中建立的。进化理论还认为生命的基本原理适用于所有的生命形式。

孟德尔的发现

孟德尔（Gregor Mendel）的实验研究的是不同豌豆品系间进行育种实验（遗传杂交）的结果，这些豌豆品系间具有不同的明确界定的特征，如种子的形状（圆滑的或皱缩的）、颜色（黄色的或绿色的）、豆荚的形状（饱满的或皱缩的）以及茎秆的长度（高秆的或矮秆的）等。他集中研究明确界定的特征间的差异，这一点非常重要。以前许多育种家曾尝试研究整体性状的遗传，如植株总重，但是没能发现任何亲本和子代间性状传递的简单的规律（框 1-1）。

框 1-1　孟德尔定律

从一个细胞世代向下一个细胞世代传递可遗传的特征的能力，是生命细胞最重要的特征。人类早已注意到遗传的存在，他们观察到许多性状，如眼睛或头发的颜色，可以由父母传递给孩子。但是，直到 20 世纪的前几年，其中的物理学基础并不清楚，遗传的染色体理论就是在那个富于创造性的时期建立的。

1860 年，人们已经知道遗传物质可以通过精子和卵子传递。1868 年 Ernst Haeckel 注意到精子主要由核物质组成，所以推测核是负责遗传的。此后，大约又花了 20 年的时间才确定了染色体是活性因子，因为有丝分裂、减数分裂和受精等细节是首先需要研究清楚的。当这些工作都完成后，可以看出，不像其他的细胞组分，染色体在子代细胞中是完全等量均分的。而且，很容易理解形成单倍体的减数分裂过程使精子和卵子染色体数目减小，因为这对保证后代染色体数目的恒定来说是必需的。但是这些事实仅仅提示染色体可能携带遗传物质。

随着遗传的基本规律的发现，在世纪交替之际获得了证据。1865 年，孟德尔在他送给位于 Brno 的自然科学学会的文章《植物杂交实验》首次提出相关的概念。在他所展示的数据中，孟德尔详细地描述了豌豆性状的传递模式（我们将在下文详细介绍）、他的关于遗传原理的结论，以及与有争议的进化论的关系等。尽管孟德尔早期做过一些努力，试图引起当时主流生物学家的兴趣，但并没有得到科学界的认可，他的观点被完全忽视了。1900 年，孟德尔去世后 16 年，三位独立工作于不同系统的植物育种者验证了被遗忘的具有重要意义的孟德尔的工作。Hugo De Vries、Karl Correns 和 Erich Tschermak 从事的是与孟德尔相关的工作，在不知道孟德尔工作的情况下分别获得了相似的结论。

独立分离律

当证实了所用的亲本是可以真实遗传的，也就是说它们会产生完全相同的后代之后，孟德尔做了一系列的只有单一性状不同（如种子形状和颜色）的豌豆亲本（P）间的杂交。所有的子代（F_1 即第一子代）只具有一个亲本的性状。例如，在黄色和绿色豌豆杂交的组合中，所有的子代都是黄色种子。在 F_1 代表现的性状称为显性（dominant），而在 F_1 代不表现的性状称为隐性（recessive）。

当孟德尔将 F_1 进行遗传杂交时，其实验结果的意义就很清楚了。这些杂交组合的结果是：大约 25%的 F_2 代个体重新呈现隐性性状，而显性性状个体大约占 75%。他所研究的 7 个性状中，每一个性状在 F_2 中的显隐性个体之比都是 3∶1。当这些实验进行到第三代（F_3）时，所有表现隐性性状的豌豆呈现真实遗传（bred true）（即产生的后代仍表现隐性性状）。而那些表现显性性状的豌豆会产生两种后代，其中 1/3 呈真实遗传（只产生显性性状后代），其余的 2/3 同样产生混合的显隐性比率为 3∶1 的后代群体。

孟德尔正确地将他的实验结果解释为(图 1-1)：各种性状是由成对的因子［现在我们称为基因（gene）］控制的，其中一个因子来自雄性亲本，另一个来自雌性亲本。例如，纯系的圆豌豆有两个拷贝［或者称为等位基因（allele）］的圆粒基因（*RR*），而纯的皱粒品系有两个拷贝的皱粒（*rr*）等位基因。每一个圆粒品系的配子带有一个圆粒基因（*R*），皱粒品系的配子带有一个皱粒基因（*r*）。在 *RR* 和 *rr* 的杂交中，受精后产生带有两种等位基因（*Rr*）的 F_1 植株，其种子表现为圆粒，因为 *R* 对 *r* 是显性的。我们把一个个体的外观或物理性状称之为表现型或表型（phenotype），而其遗传组成称为基因型（genotype），表型相同的个体可能具有不同的基因型。所以，为了了解一个生物的基因型，经常有必要进行几个世代的杂交。纯合（homozygous）一词表述的是基因对中分别来自母本和父本的基因是相同的（如 *RR* 或 *rr*）。相应地，杂合（heterozygous）一词表述的是基因对中来自母本和父本的基因是不同的（如 *Rr*）。

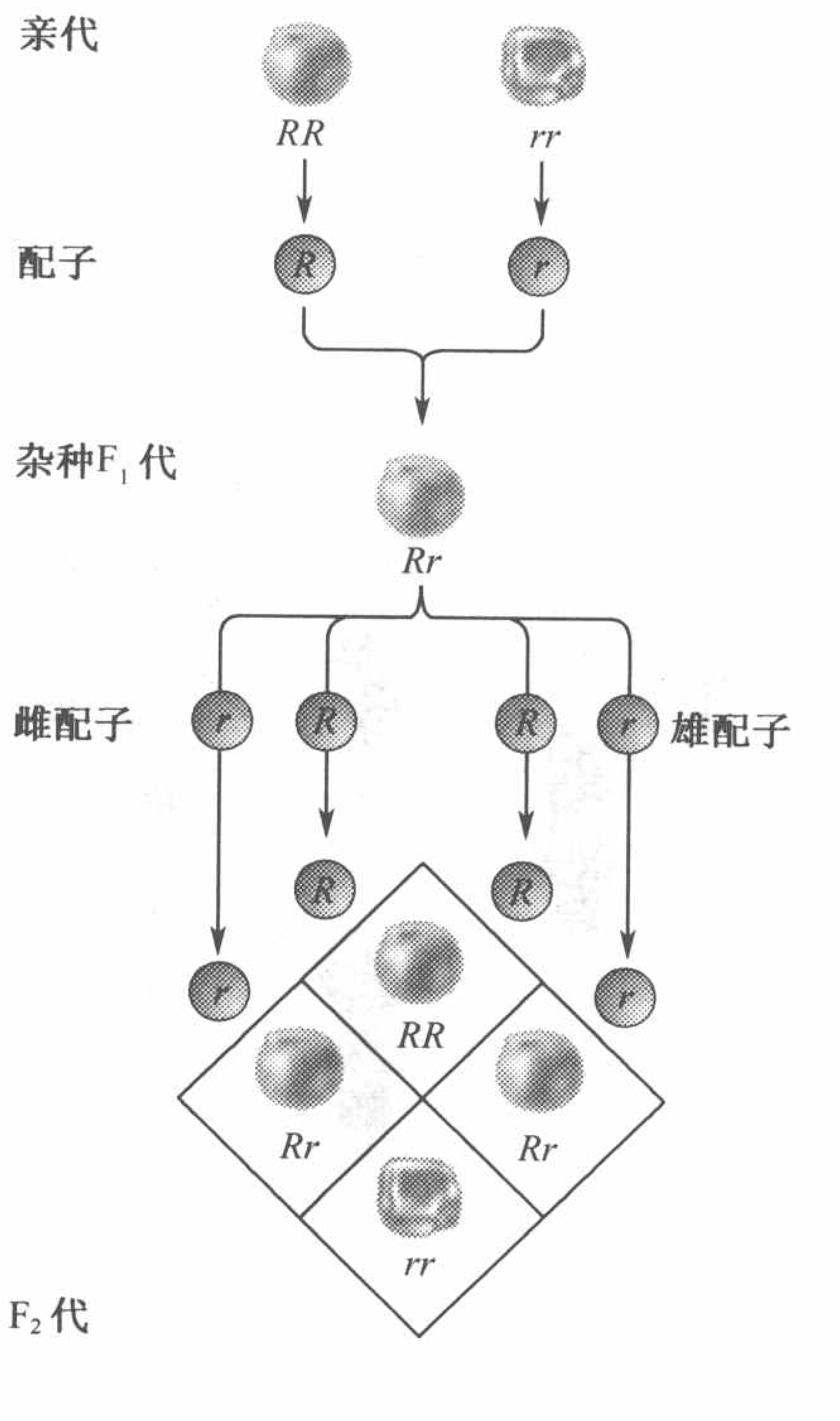


图 1-1　孟德尔第一定律（独立分离律）如何解释 F_2 代的显隐性的比率为 3∶1。 *R* 代表显性等位基因，*r* 代表隐性等位基因，圆形种子代表显性表型，皱缩的种子代表隐性表型。

可以用一个或几个字母来表示一个基因。显性等位基因可以用大写字母（*R*）或加上标（r^+）或者单独的加号（+）表示。在本书中，我们用第一种方式即大写字母表示显性的等位基因，用小写字母表示隐性的等位基因。

需要注意的是，配子中只有该生物所具有的

两个拷贝的基因中的一个（一个等位基因，要么 R，要么 r，但不会同时有两个），生物所产生的两种类型的配子数目是相同的。所以，在豌豆 F_1 代产生的配子中含有 R 或 r 的概率各占 50%，配子中带有何种等位基因完全是随机的。当我们考察数目有限的 F_2 代个体时，我们并不期望绝对的 3∶1 比率，这个比率有时稍微高一些，有时稍微低一些。但当我们观察的样品越来越多时，得到的显隐性状的个体数之比就越来越接近 3∶1。

F_2 代重新出现隐性性状，表明隐性等位基因在 F_1 代（Rr）既没有改变也没有丢失，而显性和隐性基因都是独立传递的，并且在性细胞形成过程中能够独立分离。这个独立分离律（principle of independent segregation）通常被称为孟德尔第一定律。

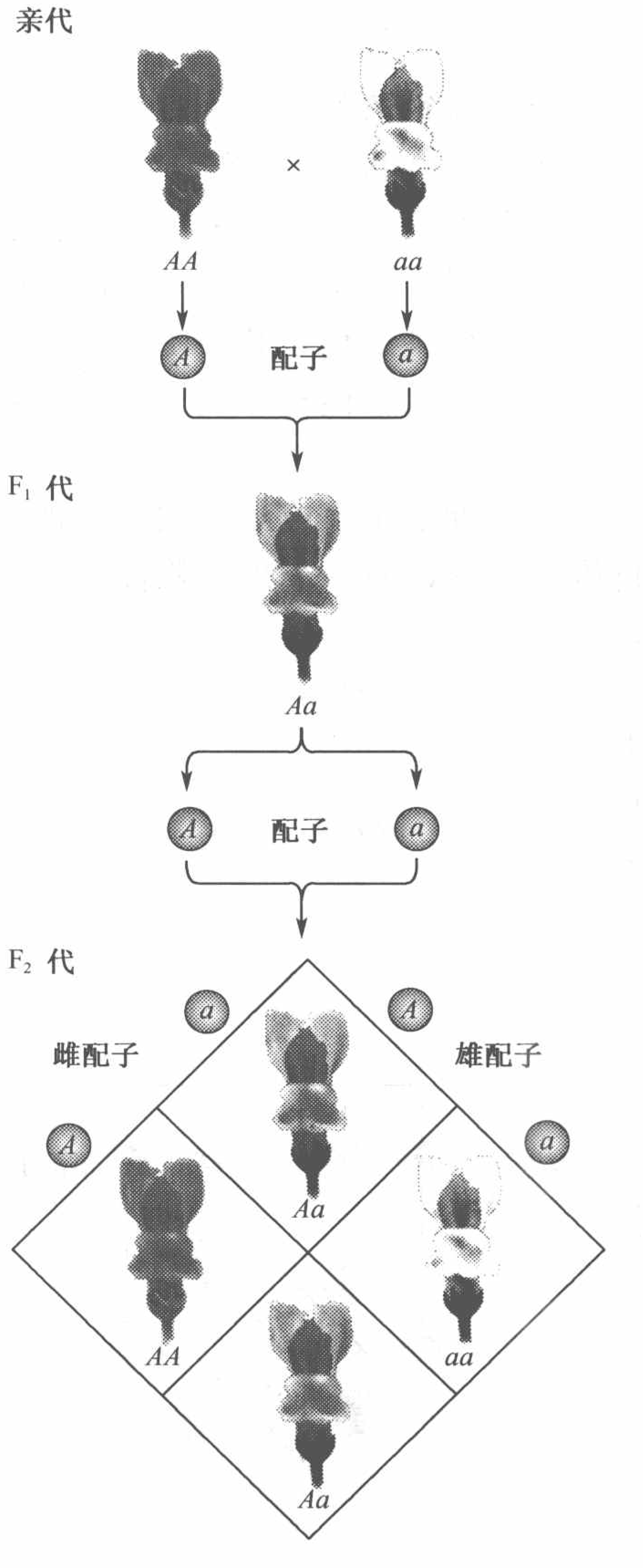


图 1-2 金鱼草花色的遗传。一个亲本是纯合的红色花（AA），另外一个亲本是纯合的白色花（aa），它们没有显隐性关系，杂合的 F_1 代花是粉红色的。F_2 代红色、粉色和白色花的比率为 1∶2∶1。

有些等位基因既非显性也非隐性

在孟德尔所报道的杂交中，基因对的一个成员对另一个成员是完全显性。但在自然界中并不总是如此。有时，杂合体的表型居于两种纯合体的表型之间。例如，纯系的红色金鱼草（*Antirrhinum*）和白色金鱼草杂交产生的 F_1 代是居间的粉红色。如果将 F_1 代自交，产生的 F_2 代个体中，红色、粉色和白色的比率为 1∶2∶1（图 1-2）。所以，在这种情况下，可以通过表型将杂合体与纯合体分开。我们也可以看到，孟德尔定律并不取决于是否基因对中的一个基因对另一个基因呈显性。

独立分配律

孟德尔将他的育种实验扩大到有多个不同性状的豌豆品系之间。就像上文一样，他从两个豌豆品系开始，自交时它们都是真实遗传的。一个品系具有圆形的黄色种子，另一个是皱缩的绿色种子，因为圆形黄色对皱缩绿色是显性的，所有的 F_1 代种子都是圆形黄色的。将 F_1 代自交，检查所产生的 F_2 代种子的外观（表型），除两种原来的表型外（圆形黄色、皱缩绿色），又出现了两种新的表型［重组体（recombinant）］：皱缩黄色和圆形绿色。

同样地，孟德尔发现如果假定基因存在，并且每一对基因在性细胞形成过程中，都可以独立地传递到配子中，他就可以解释这些

结果。这种解释如图 1-3 所示，每一个配子只包含基因对中的一个等位基因。所以，F_1 代（*RrYy*）产生的配子将是 *RY*、*Ry*、*rY*、*ry*，但不会是 *Rr*、*Yy*、*YY* 或 *RR*。进一步讲，在这个例子中，出现 4 种可能的配子的频率都是相同的，来自同一亲本的基因并没有同时出现的倾向。那么结果就是，F_2 代的表型就呈现 9 个圆形黄色、3 个圆形绿色、3 个皱缩黄色和 1 个皱缩绿色，就像图 1-3 的下部的 Punnet 方块所展示的（这种排列方式由英国数学家 Punnet 介绍的，所以以他的名字命名）。这一独立分配律（principle of independent assortment）通常被称作孟德尔第二定律。

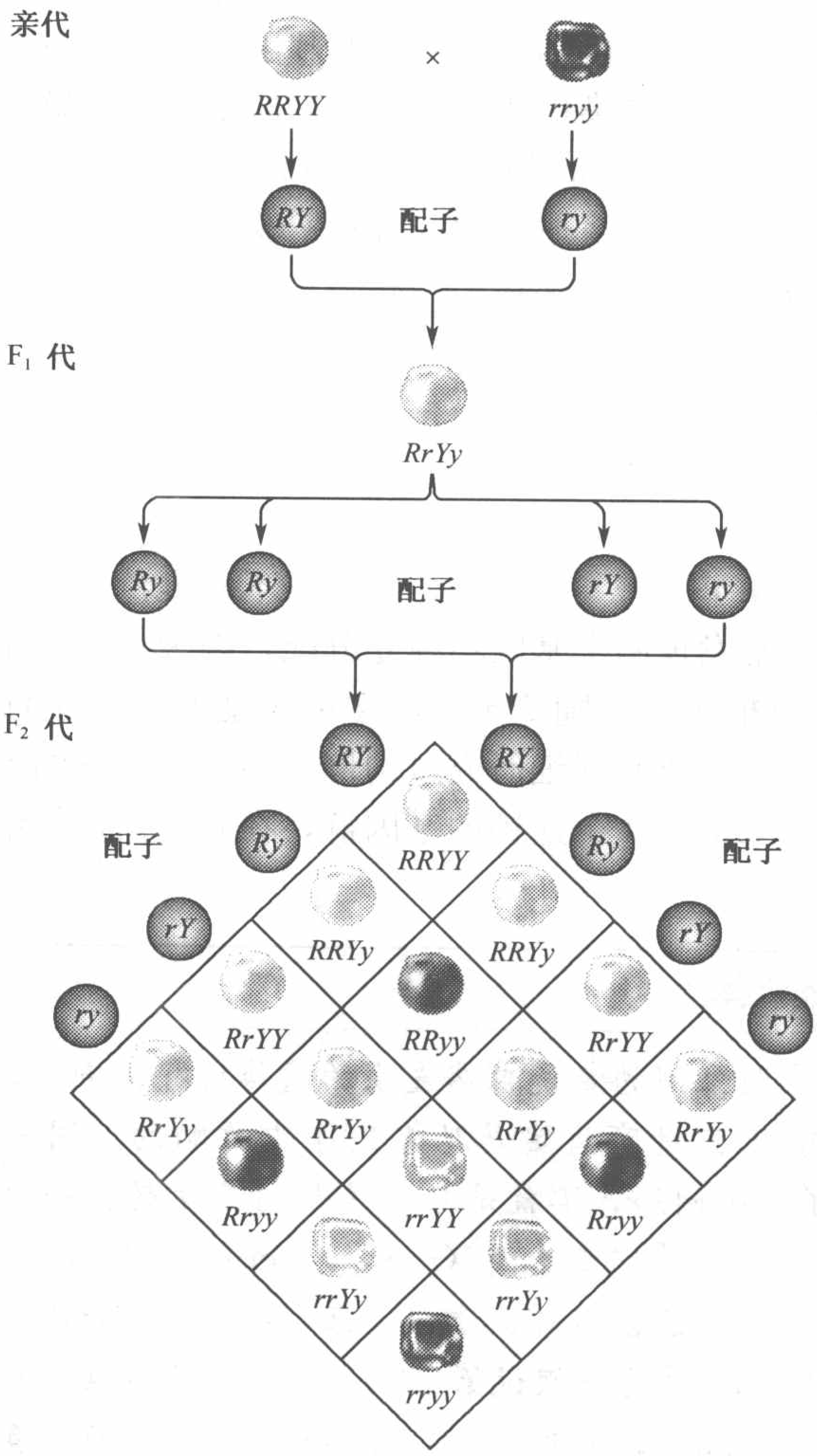


图 1-3　孟德尔第二定律（独立分配律）是如何发挥作用的。在这个例子中，黄色（*Y*）和绿色（*y*）种子的遗传与圆形（*R*）和皱缩（*r*）种子的遗传是连在一起的。*R* 和 *Y* 等位基因对 *r* 和 *y* 呈显性。亲本和子代的基因型用相应的字母组合表示，4 种不同的表型用相应的背景色来区分。

遗传的染色体理论

孟德尔的发现开始不被认可的主要原因就是缺乏减数分裂和有丝分裂期间染色体行为事实的支持。但到了1900年，当孟德尔规律被重新发现时，人们已经有了这方面的知识。1903年美国生物学家Walter S. Sutton在他的经典文章“遗传过程中的染色体”一文中，发现二倍体由两组形态上相似的染色体组成，在减数分裂过程中，每一个配子都获得一对同源染色体中的一条，他假设基因是染色体上的一部分，并用上述发现来解释孟德尔的实验结果。他推测黄色和绿色种子的基因是在某一特定染色体对上，而圆粒和皱缩粒基因则是在另一不同的染色体对上。这一假说立即就解释了实验中所观察到的9∶3∶3∶1的分离比。虽然Sutton的文章并没有证明遗传的染色体理论，但它的意义还是非常重大的，因为他第一次将遗传学（育种实验的研究）和细胞学（细胞结构的研究）这两个独立学科的实验结果整合在了一起。

基因连锁和交换

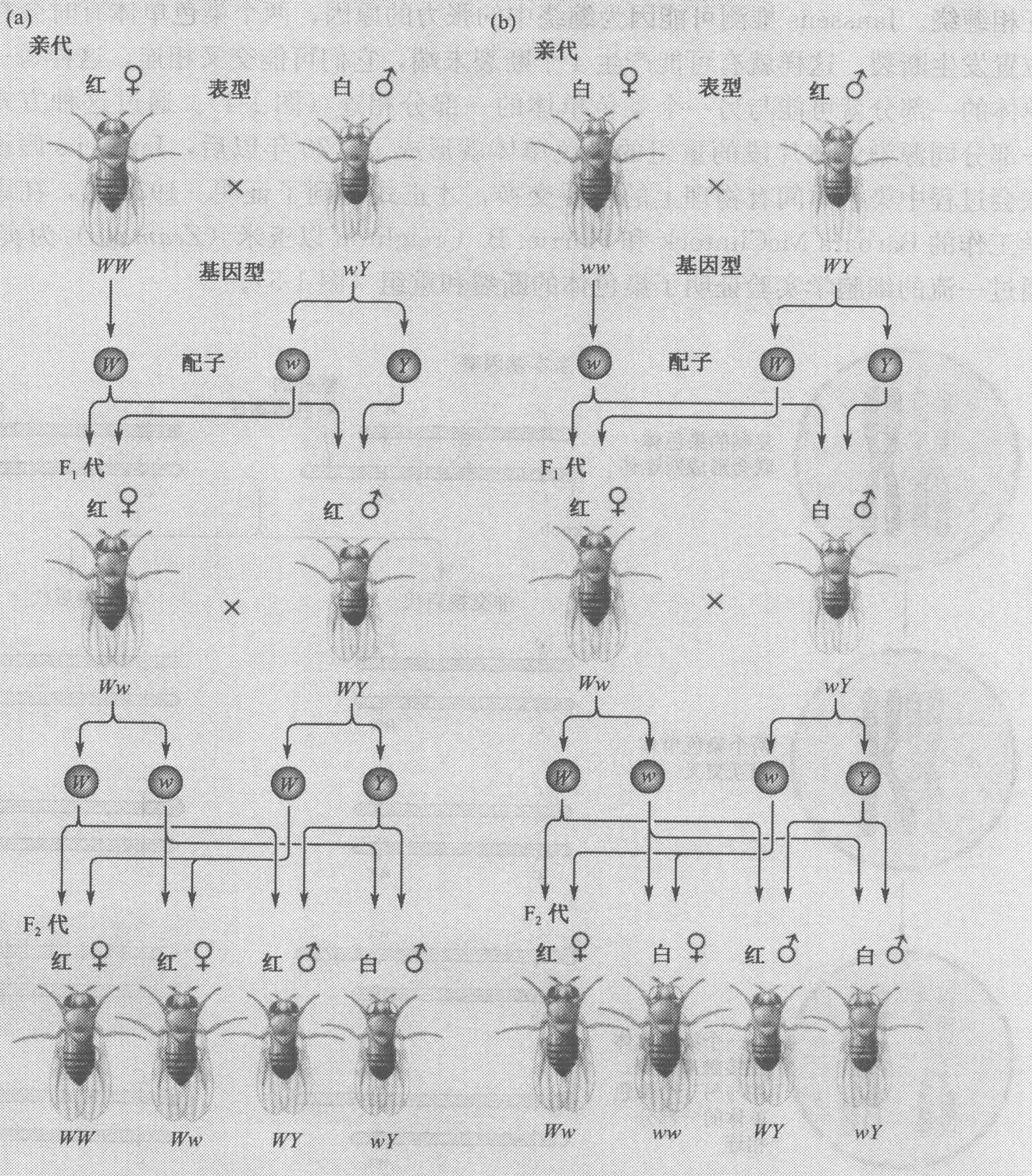
孟德尔的独立分配律是基于一个事实，即位于不同染色体上的基因在减数分裂过程中独立行动。然而，经常发现两个基因不独立分配，因为它们位于同一染色体上[连锁基因（linked gene），见框1-2]。随着越来越多的突变基因用于育种分析，许多不符合随机分配的例子被发现。在所有仔细研究过的例子中，连锁群的数目与单倍体染色体的数目是相同的。例如，果蝇有4个连锁的基因群，在其单倍体体细胞中有4个形态上可分的染色体。

框1-2 基因串联在染色体上

最初所有的育种实验采用的都是自然发生的遗传变异。例如，孟德尔使用过从种子经营商处得来的种子，而经营商肯定是从农民手中收购的。同一基因存在不同的形式（等位基因）就产生了一个问题，那就是它们是如何产生的？一个直接的假说就是基因可以变为（突变为）新基因[突变基因（mutant gene）]。从1908年开始，伟大的美国生物学家Morgan和他的年轻合作者，遗传学家Calvin B. Bridges、Hermann J. Muller以及Alfred H. Sturtevant对这一假说第一次进行了认真的实验验证。他们以小小的果蝇（*Drosophila melanogaster*，黑腹果蝇）为实验材料，发现了第一个雄性白眼果蝇突变体，而不是正常的红眼。白眼突变是在红眼果蝇的培养瓶中自发产生的突变体。实际上，自然界中存在的几乎所有果蝇都是红眼的，控制红眼的基因就称为野生型基因（wild-type gene），导致白眼的基因称为突变基因（等位基因）。

白眼突变基因马上被用来进行育种实验（框1-2图1），实验结果令人兴奋，因为等

位基因的表现与X染色体完全平行（也就是说，与性是连锁的）。这一发现意味着这个基因可能与控制性别的基因都位于X染色体上。这一假说很快就由更多的新突变基因的交配实验得到了验证，当时新发现的基因中多数也是性连锁的。



框1-2图1　果蝇中的性连锁基因的遗传。位于性染色体上的基因会分别在雄性和雌性后代表现，因为如果只有一条X染色体，该染色体上的隐性基因总是会表现。图示两个交配实验，都涉及位于X染色体上的隐性基因[*w*代表白眼（white eye）]。(a) 雄性亲本是白眼果蝇（*wY*），雌性是同源的红眼个体（*WW*）；(b) 雄性为红眼（*WY*），雌性为白眼（*ww*）。此处字母Y不是代表的等位基因，而是雄性果蝇中对应于X染色体的Y染色体。Y染色体上没有对应于X染色体上的*w*或*W*的基因。

然而，实际上连锁不可能是完全的。在减数分裂过程中，同一染色体上的两个基因保持同步的概率，其范围从略低于100%到接近50%。这一数值的变化表明存在着一个同源染色体间基因交换的机制，这个机制就称为交换（crossing over）。比利时细胞学

家 F. A. Janssens 首先描述了其细胞学基础。在减数分裂开始时，通过一个联会过程(synapsis)，同源染色体沿着与长轴平行的方式形成染色体对。在这个阶段，每一个染色体都复制一次，形成两个染色单体。所以，联会将 4 个染色单体聚在一起（四连体），互相缠绕。Janssens 推测可能因为缠绕中的张力的原因，两个染色单体有时会在相应的位置发生断裂。这样就有可能产生 4 个断裂末端，它们可能交叉相连，这样每一个染色单体的一部分就可能与另一个染色单体的一部分相连（图 1-4）。通过这种方式，含有一部分同源染色体片段的重组的染色单体就形成了。20 年以后，Janssens 假说，即在联会过程中染色体间有物理上的物质交换，才正式得到了证明。1931 年，在康奈尔大学工作的 Barbara McClintock 和 Harriet B. Creighton 以玉米（*Zeamays*）为实验材料，通过一流的细胞学实验证明了染色体的断裂和重组（图 1-5）。

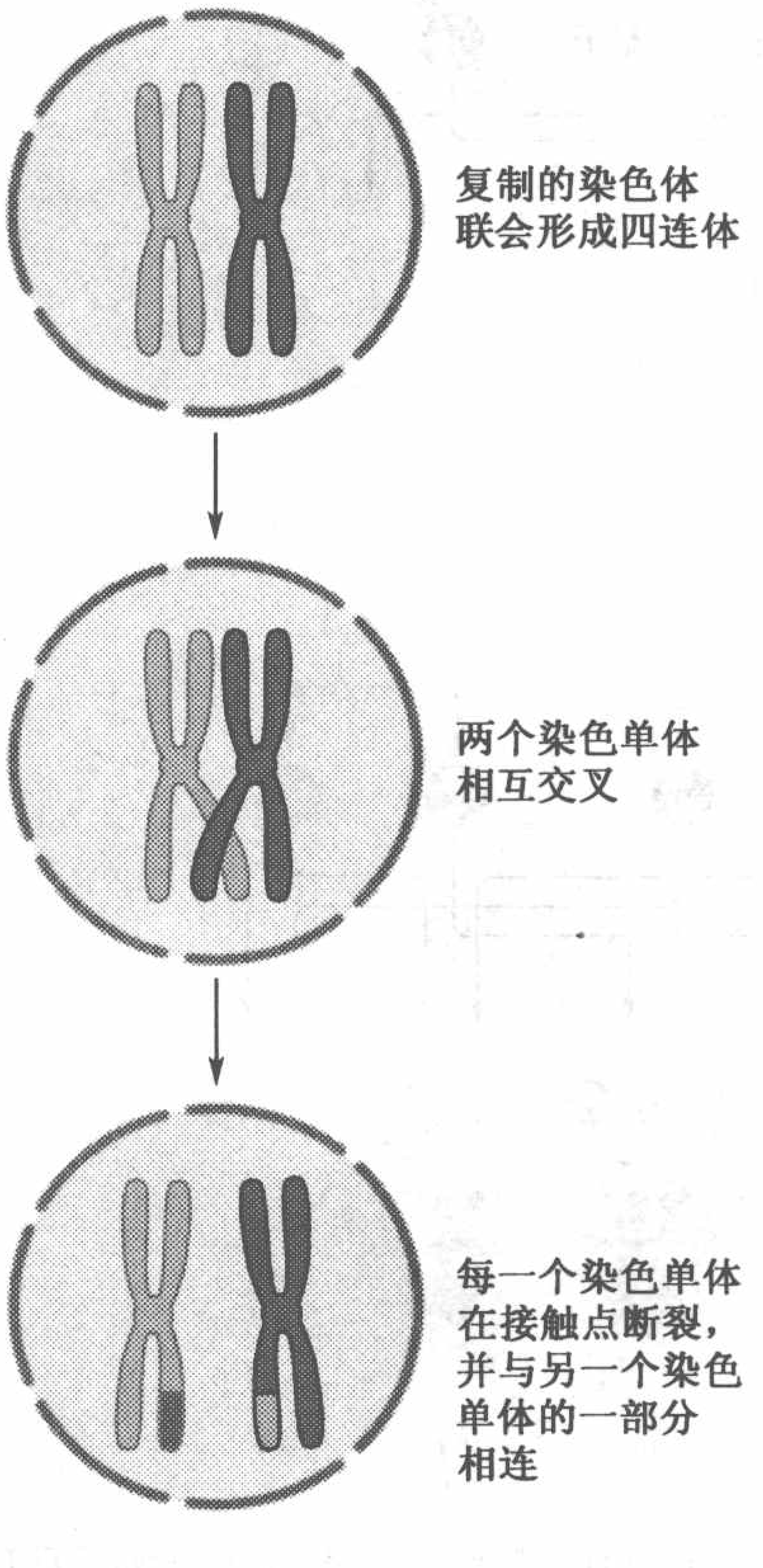


图 1-4 Janssens 的交换假说。复制的染色体联会形成四连体。

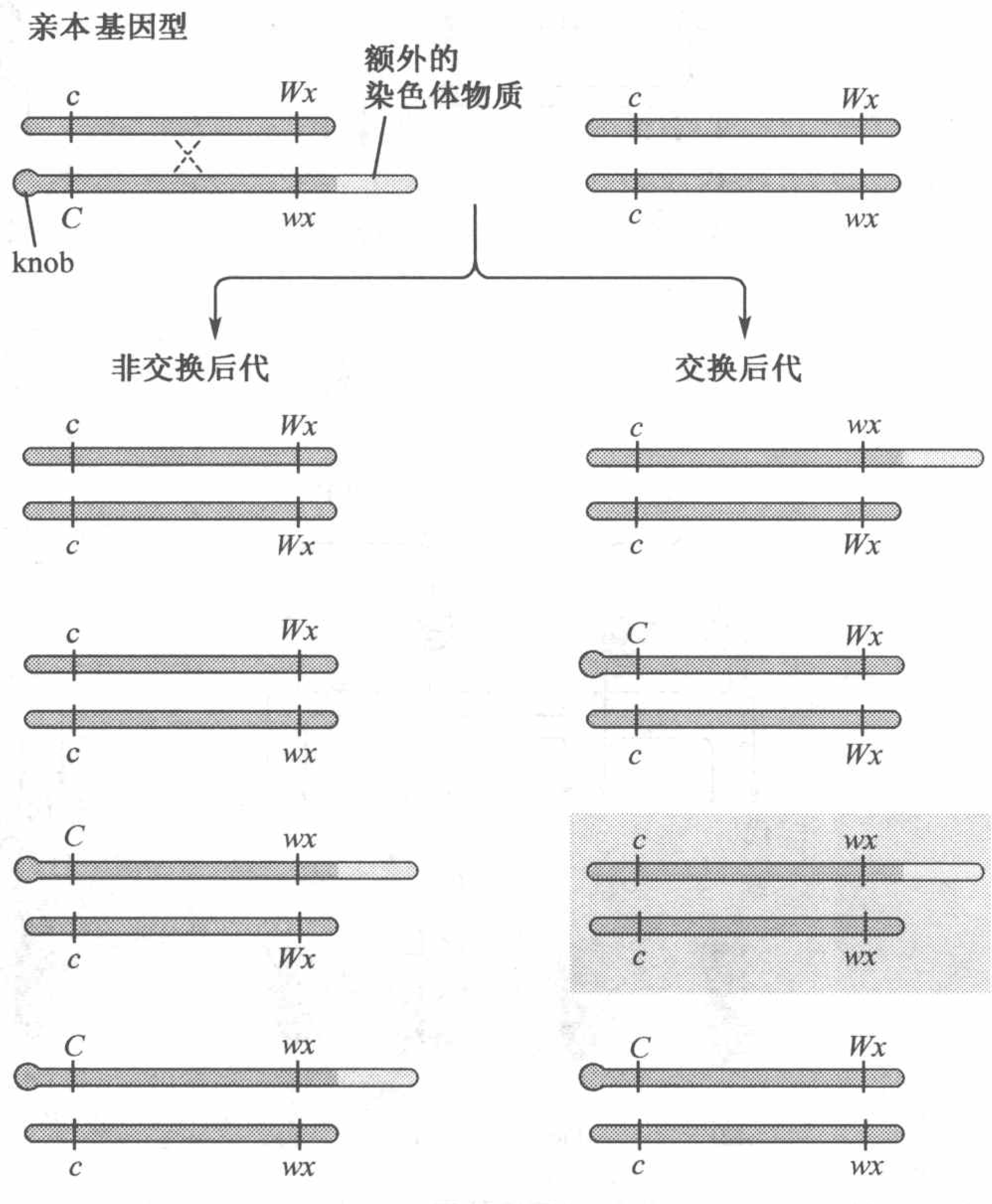


图 1-5 同源染色体间物理交换的示意图。在大多数生物中，同源的染色体对具有相同的形状。然而，有时它们也不完全相同，图中所示一条染色体带有额外的一段遗传物质，以及一段致密的可以反复形成节状结构的区域。McClintock 和 Creighton 就发现了这样一对染色体，并用它们证明在配对的染色体间涉及实际的物理交换。在图示的实验中，同源的 *c*，*wx* 后代肯定是在 *C* 和 *wx* 基因座间发生交换后产生的。当用细胞学方法检测 *c*，*wx* 后代时，可以看到具有节状结构的染色体，表明无节的 *Wx* 区域是被有节的 *wx* 区域物理上替代了。图中带阴影的框标示的是纯合的 *c*，*wx* 子代染色体。

染色体定位

摩尔根和他的学生们，并没有等待交换发生的细胞学证据就着手探索 Janssens 假说的含义。他们推测染色体上紧密相邻的基因（紧密连锁）比相距较远的基因更可能会同时分配。他们马上就意识到，这是一种确定染色体上的基因相对位置，进而绘制遗传图谱（genetic map）的方法，他们利用的不同重组群发生频率的方法是很直观的。让我们来看同一个染色体上 3 个基因分离的情形，这些基因的排列状况可以用 3 次交配来确定，每次交配只调查其中的两个基因（二因子交配）。*AB* 和 *ab* 之间的交配会产生 4 种类型的后代：两种亲本基因型（*AB* 和 *ab*）、两种重组基因型（*Ab* 和 *aB*）。*AC* 和 *ac* 之间的交配同样会产生两种亲本基因型和两种重组基因型 *Ac* 和 *aC*，而 *BC* 和 *bc* 之间的交配将产生亲本基因型和 *Bc*、*bC* 重组基因型。每一个交配都将产生一个特定的亲本基因型和重组基因型后代的比率。例如，第 1 组交配产生 30%的重组子，第 2 组产生 10%，第 3 组 25%。这些数据告诉我们，基因 *a* 和 *c* 的距离较 *a* 和 *b* 以及 *b* 和 *c* 的距离近，而 a 和 b 之间与 b 和 c 之间的遗传距离更为接近，最符合这些数据的基因排列顺序是 *a-c-b*（图 1-6）。

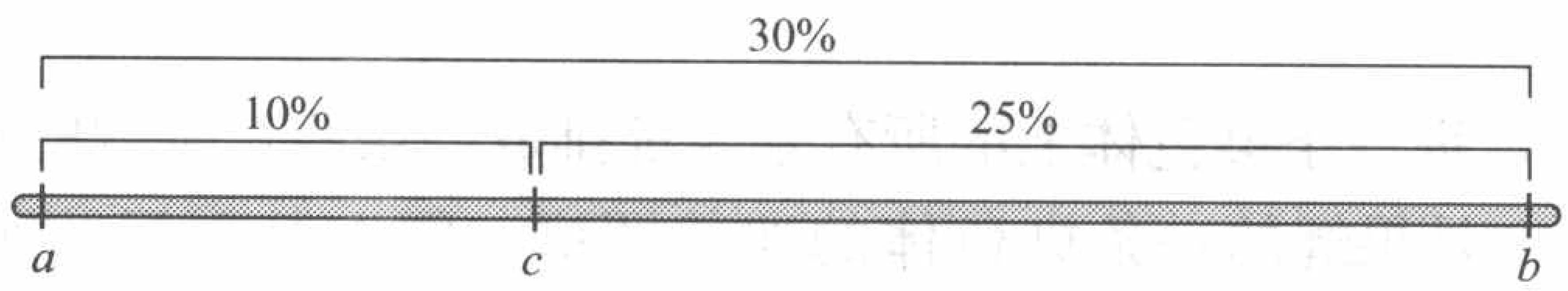


图 1-6　根据 3 个二因子实验，推测 3 个基因的顺序。

通常会用三因子交配实验来准确无误地验证二因子交配实验所确定的基因顺序的正确性。当前述例子中的 3 个基因通过 *ABC*×*abc* 的交配进行时，会出现 6 种重组基因型（图 1-7），它们可分为 3 组，个体数目最少的一组来自双交换。通过寻找频率最低的那一组，往往立即就可以确认（或否定）一种推测的排列方式。图 1-7 的结果马上就可以确认二因子交配实验所暗示的顺序，事实上只有 *a-c-b* 顺序才是合理的，即最少的重组子是 *AcB* 和 *aCb*。

存在多个交换事件的事实意味着最远的两个标记 *a* 和 *b*（*ab*）交换发生重组的数目通常少于 *a* 和 *c*（*ac*）以及 *c* 和 *b*（*cb*）之间重组概率的和。为了得到更精确的最远标记的距离，我们来计算 *ac*×*cb*（重组的）概率，即在 *c* 和 *b* 间发生一次交换，在 *a* 和 *c* 间也发生一次交换，或者是反过来（*cb*×*ac*）。将这个概率从总的概率中减去，就会得到更准确的重组的概率。如下的简单公式

$$ab = ac + cb - 2(ac)(cb)$$

适合于一个交换的发生不影响其他交换发生频率的所有情形。但遗憾的是，精确的定位经常受干涉现象（interference）的扰乱，它会提高或降低相关交换的概率。

利用上述推理，由摩尔根领导的哥伦比亚大学研究小组在 1915 年之前确定了多于 85 个果蝇突变基因的位置（表 1-1）。每个基因被放在 4 个连锁群，或者说染色体的一个特定

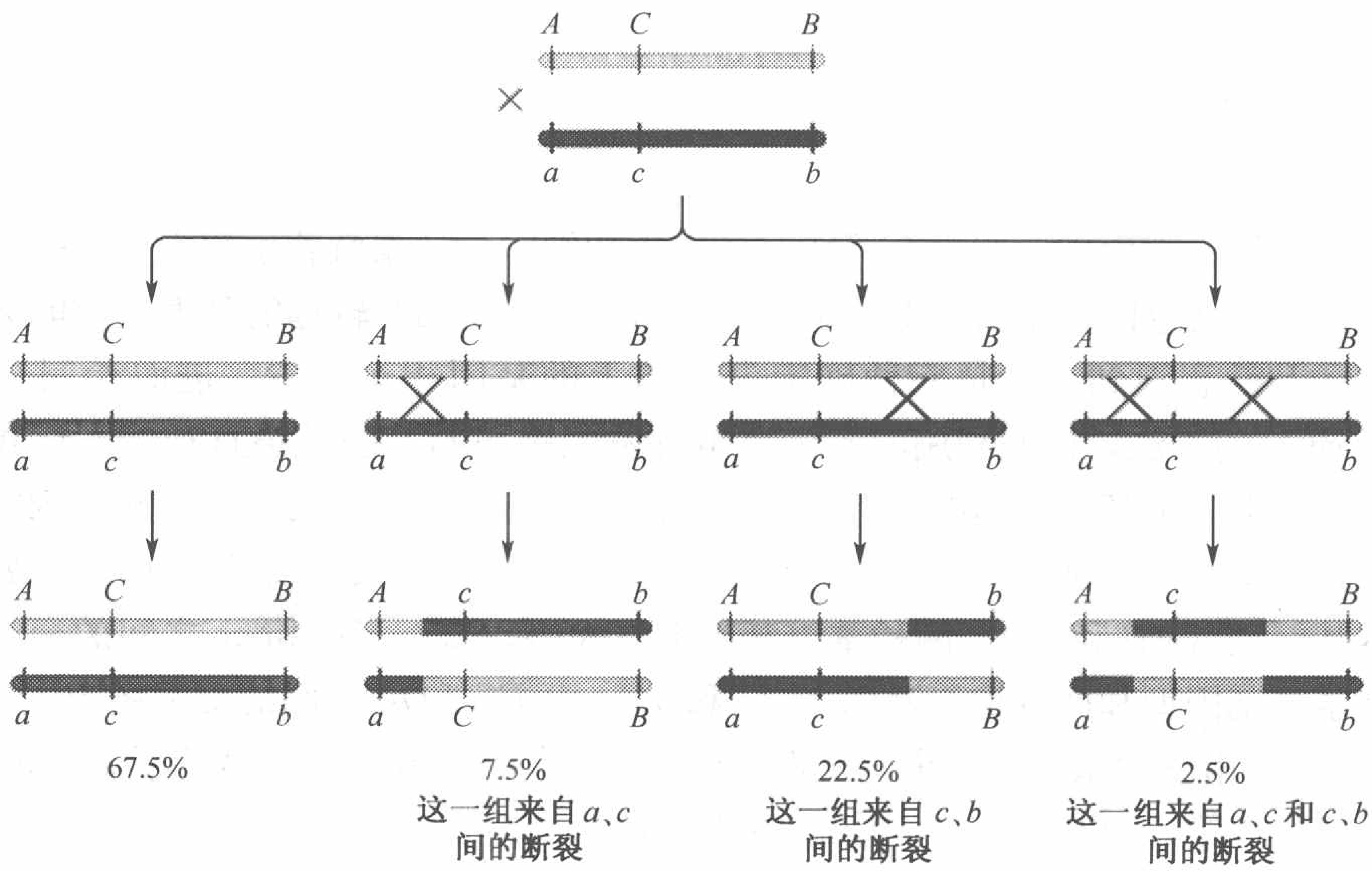


图 1-7　用三因子交配确定基因的顺序。发生频率最低的交互重组一定是来自一个双交换，图中所列的不同的百分比是从无限多的样品中获得的理论值。当只有一定数目的后代时，其准确值的确定需要考虑随机的统计。

位点。最重要的是，每一条染色体上的所有基因都排成一条线，基因的排列是严格线性而没有分支的。图 1-8 是果蝇的一条染色体的遗传图谱，图上基因间的距离是用图距单位（map unit）表示的，图距单位是与基因间的重组频率相关的。如果两个基因间的重组频率是 5%，那么基因间的距离就是 5 个图距单位。因为在相距很远的基因间发生双交换的概率很高，图距单位的计算就不准确，除非所有紧密相邻的基因间都进行了重组实验。

表 1-1　1915 年报道的 85 个果蝇突变体基因*

名称	所影响的位置	名称	所影响的位置
第 1 组			
Abnormal	腹部	*Lethal*，13	身体，死亡
Bar	眼睛	*Minature*	翅膀
Bifid	脉络	*Notch*	脉络
Bow	翅膀	*Reduplicated*	眼睛颜色
Cherry	眼睛颜色	*Ruby*	腿
Chrome	身体颜色	*Rudimentary*	翅膀
Cleft	脉络	*Sable*	身体颜色
Club	翅膀	*Shifted*	脉络
Depressed	翅膀	*Short*	翅膀
Dotted	胸部	*Skee*	翅膀
Eosin	眼睛颜色	*Spoon*	翅膀
Facet	小眼	*Spot*	身体颜色
Forked	脊骨	*Tan*	触角
Furrowed	眼睛	*Truncate*	翅膀
Fused	脉络	*Vermilion*	眼睛颜色
Green	身体颜色	*White*	眼睛颜色
Jaunty	翅膀	*Yellow*	身体颜色
Lemon	身体颜色		

续表

名称	所影响的位置	名称	所影响的位置
第 2 组			
Antlered	翅膀	*Jaunty*	翅膀
Apterous	翅膀	*Limited*	腹部条纹
Arc	翅膀	*Little crossover*	2 号染色体
Balloon	脉络	*Morula*	小眼
Black	身体颜色	*Olive*	身体颜色
Blistered	翅膀	*Plexus*	脉络
Comma	胸部标记	*Purple*	眼睛颜色
Confluent	脉络	*Speck*	胸部标记
Cream Ⅱ	眼睛颜色	*Strap*	翅膀
Curved	翅膀	*Streak*	体型
Dachs	腿	*Trefoil*	体型
Extra vein	脉络	*Truncate*	翅膀
Fringed	翅膀	*Vestigial*	翅膀
第 3 组			
Band	体型	*Pink*	眼睛颜色
Beaded	翅膀	*Rough*	眼睛
Cream Ⅲ	眼睛颜色	*Safranin*	眼睛颜色
Deformed	眼睛	*Sepia*	眼睛颜色
Dwarf	身体大小	*Sooty*	身体颜色
Ebony	身体颜色	*Spineless*	脊骨
Giant	身体大小	*Spread*	翅膀
Kidney	眼睛	*Trident*	体型
Low crossing over	3 号染色体	*Truncate*	翅膀
Maroon	眼睛颜色	*Whitehead*	体型
Peach	眼睛颜色	*White ocelli*	单眼
第 4 组			
Bent	翅膀	*Eyeless*	眼睛

* 突变发生于 4 个连锁群中，因为细胞学研究中观察到的染色体数目为 4，表明基因是位于染色体上。注意不同基因的突变会对单一的性状（如身体的颜色）有不同的影响。

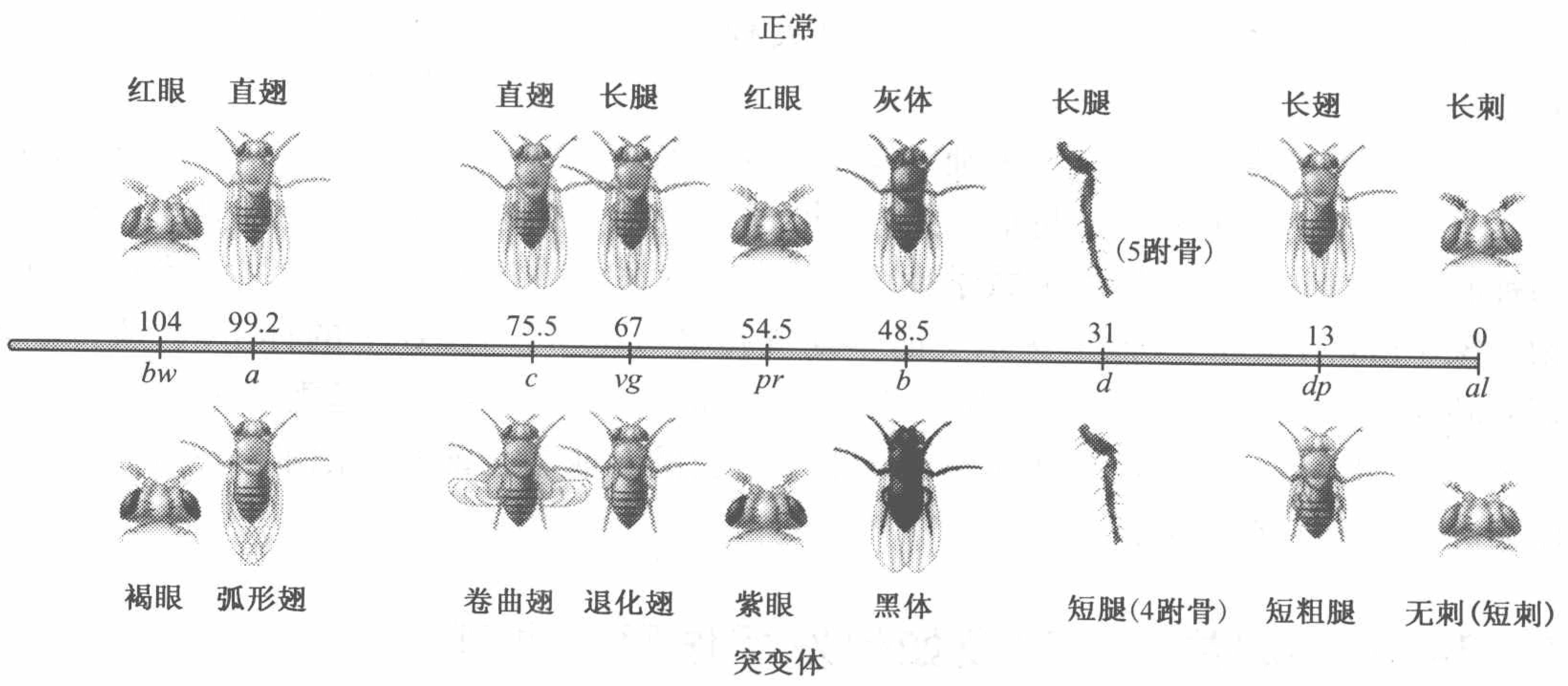


图 1-8　果蝇的第 2 号染色体的遗传图。

即使两个基因处在很长的染色体的两端，因为有多个交换事件发生，它们至少有

50%的概率共同分离。如果它们中间发生了奇数次的交换，两个基因就会分开。如果发生的交换次数为偶数，它们就不会分开。所以，在果蝇遗传研究的早期，经常不能确定两个基因是位于不同的染色体上还是位于一条长染色体的两端。只有定位了很多的基因后，才能令人信服地证明连锁群的数目与细胞学上可分的染色体数目相等。1915 年，摩尔根与他的学生 Alfred H. Sturtevant、Hermann J. Muller 和 Calvin B. Bridges 发表了权威性的著作《孟德尔遗传的机制》，在这本书中，他们首次宣布了遗传的染色体基础的普遍性。现在我们将这一概念与进化论和细胞理论一起作为理解生命本质的最重要成就。

突变引起遗传变异

现在我们有可能理解生物界中那些可遗传的变化是进化理论的基石。正常情况下，染色体复制过程中基因是忠实复制的。然而在极少数情况下，基因中还是会发生一些变化［突变（mutation）］，形成不同的等位基因，其中大部分（当然不是全部）的功能不如野生型的等位基因。显然，这样的事情应该很少发生，否则，在每一个细胞周期中都有很多基因发生突变，那么产生的后代就不再像它们的亲本。实际上，存在少量而有限度的突变是有很多好处的，它是一个产生变异体的稳定来源，是动植物适应不断变化的物理和生物环境所需要的。

然而，出乎意料的是，这些孟德尔学派遗传学家们的实验结果并没有被传统的生物学家所接受，研究不同生命形式的进化关系的权威人士也不接受。人们的疑问是，摩尔根和他的学生们所研究的这些遗传变异是否足以产生进化上明显不同的结构，如翅膀和眼睛。相反，这些生物学家相信，肯定还有更强有力的“大突变”存在，它们才是进化跃进得以发生的原因。

然而慢慢地疑问消失了，其中在很大程度上应该归功于数学遗传学家 Sewall Wrigh、Ronald A. Fisher 和 John Burden Sanderson Haldane 等的工作。他们的工作表明，考虑到地球漫长的岁月，果蝇中发现的相对较低的突变率，以及温和的选择优势，应该足够积累起新的有益的特征。到 20 世纪 30 年代，生物学家开始重新评价他们关于物种起源的知识，并开始理解数学遗传学家们的工作。在这些新达尔文主义者中，有生物学家 Julian Huxley（他是达尔文的早期支持者 Thomas Huxley 的孙子）、遗传学家 Theodosius Dobzhansky、古生物学家 George Gaylord Simpson 和鸟类学家 Ernst Mayr 等。在 20 世纪 40 年代，上述 4 位都著书从自己的专业角度证明孟德尔学派和达尔文学派实际上是相互符合的。

早期关于基因是什么以及如何发挥作用的推测

几乎就在孟德尔定律重新发现的同时，遗传学家就开始推测基因的化学结构以及它们是如何发挥功能的。因为无从得知遗传物质的化学特性，所以一直没有真正的进展。

即使当时已经知道染色体中存在核酸和蛋白质也帮助不大，因为对它们的结构都一无所知。最有价值的推测是基于一个事实，即基因肯定可以自我复制，至少在某种意义上应该如此。当一条染色体变为两份时，它们的结构肯定会得到准确的复制。这一事实立即引出了一个非常重要的化学问题，那就是一个复杂的分子是如何精确地被复制并产生与原件无异的复制子的。

一些物理学家也开始对基因感兴趣，当量子力学在 20 世纪 20 年代后期横空出世时，一种观点出现了：为了了解基因，必须首先了解最先进的理论物理的奥妙。然而，这种想法从来没有真正执行，因为很明显即使最好的物理学家或理论化学家也不会让自己关心结构尚待阐明的物质。只有一个事实也许会让他们有所思考，那就是 1927 年 Muller 和 L. J. Stadler 独立发现的 X 射线可以诱发突变产生。因为 X 射线轰击大的基因的概率比轰击小基因的概率要高，所以可以根据某一射线剂量下照射某一特定基因诱发突变的频率大致估计基因的大小。即便如此还需要种种特定的假设，事实上没有任何人，包括 Muller 和 Stadler 本人在内，都没有真正地重视这种可能性。

发现基因与蛋白质相互关系的初步尝试

检查某个基因的改变会影响细胞中的哪一个蛋白质，是发现基因和蛋白质相互关系的最有成效的早期努力。最初这些研究相当困难，因为没有人知道器官中（如眼睛或翅膀）的蛋白质的任何信息，不久人们就意识到研究简单的具有代谢功能的基因应该比研究影响整体结构的基因更容易，一个早期有效的例子来自影响某种氨基酸代谢的一种遗传病的研究。在人类中会发生影响苯丙氨酸代谢能力的自发突变，当一个带有纯合的突变性状的个体吃了含苯丙氨酸的食物时，由于他不能将苯丙氨酸转化为络氨酸，就会造成苯基喹啉酸（phenylpyruvic acid）在血液中增加至毒性水平，这些疾病称为“先天代谢错误”。早在 1909 年，英国医生 Archibald E. Garrod 就认为其野生型基因是负责某种特定酶的表达，在纯合的突变体中，这种酶将先天缺乏。

Garrod 提出的关于基因与蛋白质相互关系的一般假说在 20 世纪 30 年代通过下列工作而得到进一步拓展：Haldane 和 Rose Scott-Moncrieff 在英格兰进行的关于花色素工作，美国的 Wright 在豚鼠的毛发色素方面的工作，德国的 A. Kuhn 以及 Boris Ephrussi 和 George W. Beadle 起初在法国，后来在加州（美国）进行的关于昆虫眼睛色素的研究。在所有的研究中，实验证据都表明特定的基因影响色素形成的特定的步骤，基因的缺失就会造成颜色的改变，例如，果蝇眼睛的颜色由红色变为琥珀色。然而，由于缺乏相关酶的结构的基本知识，所以不可能进一步分析基因与酶之间的关系，并且当时也不能确认大部分基因是控制蛋白质合成的（当时人们对所有的酶都是蛋白质也是有怀疑的），也不清楚是否所有的蛋白质都是受基因控制的。

早在 1936 年，孟德尔学派的遗传学家就清楚地意识到，如果仍沿用那些阐明孟德尔遗传学的基本特征的方法进行实验，是难以有效地获得关于基因是如何发挥作用的信息的。相反，取而代之的应该是寻找更适合化学分析的生物学目标。他们认识到当时拥

有的关于核酸和蛋白质的化学知识是完全不足以来进行基础的生物化学研究的，即使是最合适的生物学系统也不行。然而，幸运的是，化学知识的局限性并没有阻止他们学习如何用化学上简单的霉类、细菌和病毒进行遗传实验。就像我们将要看到的，就在遗传学家们需要的时候，必要的化学知识也产生了。

小结

遗传是由染色体控制的，它是细胞内基因的载体。遗传因子在 1865 年首先由孟德尔发现并进行描述，但是其重要意义到 20 世纪初才被认可。每一个基因都可能存在不同的形式，称为等位基因。孟德尔认为决定一个遗传性状的遗传因子（也就是现在众所周知的基因）是由每个亲本提供给每个后代的。这种行为的物理基础是在减数分裂过程中同源染色体的分配：一对同源染色体中的一条（随机选择）将分配给一个单倍体细胞，当两个基因位于同一染色体上时，它们倾向于共同遗传（连锁）。决定不同性状的基因经常是相互独立遗传的，因为它们位于不同的染色体上的。在任何情形下，连锁极少是完全的，因为在减数分裂过程中同源染色体是相互结合的，并且经常在相同处断裂后重新交互连接（交换），交换可以将来自父本染色体的基因转移到来自母本染色体的基因群中。

来自同一基因的不同的等位基因源自基因自身的可遗传的变异（突变）。正常情况下，基因非常稳定，并且在染色体复制过程中是忠实拷贝的；突变只是偶尔发生的，一般还会产生有害的后果。然而突变扮演着一个有用的角色，有利的稀有突变的积累是进化理论所依赖的遗传多样性的基础。

多少年来，基因的结构及其控制细胞特征的化学途径是神秘的。随着大量的自发突变体的研究，人们越来越清楚，一个基因控制一个性状的关系是不存在的，所有复杂性状都受控于许多基因。最合理的观点是由 Garrod 在 1909 年提出的“基因控制酶的合成”。然而，孟德尔学派遗传学家的工具，如植物玉米、动物小鼠，甚至果蝇等生物材料都不适合用于基因与蛋白质之间的深入的化学研究。为了进行这类分析，用更简单的生物进行实验是必不可少的。

（刘国振 译 吴东颖 李雷 刘国振 刘韧 杨焕明 校）

参考文献

Ayala F.J. and Kiger J.A., Jr. 1984. *Modern genetics*, 2nd ed. Benjamin Cummings, Menlo Park, California.

Beadle G.W. and Ephrussi B. 1937. Development of eye color in *Drosophila:* Diffusible substances and their interrelations. *Genetics* **22:** 76–86.

Carlson E.A. 1966. *The gene: A critical history.* Saunders, Philadelphia.

——— 1981. *Genes, radiation, and society: The life and work of H.J. Muller.* Cornell University Press, Ithaca, New York.

Caspari E. 1948. Cytoplasmic inheritance. *Adv. Genet.* **2:** 1–66.

Correns C. 1937. *Nicht Mendelnde vererbung* (ed. F. von Wettstein). Borntraeger, Berlin.

Dobzhansky T. 1941. *Genetics and the origin of species,* 2nd ed. Columbia University Press, New York.

Fisher R.A. 1930. *The genetical theory of natural selection.* Clarendon Press, Oxford.

Garrod A.E. 1908. Inborn errors of metabolism. *Lancet* **2:** 1–7, 73–79, 142–148, 214–220.

Haldane J.B.S. 1932. *The courses of evolution.* Harper & Row, New York.

Huxley J. 1943. *Evolution: The modern synthesis.* Harper & Row, New York.

Lea D.E. 1947. *Actions of radiations on living cells.* Macmillan, New York.

Mayr E. 1942. *Systematics and the origin of species.* Columbia University Press, New York.

——— 1982. *The growth of biological thought: Diversity, evolution, and inheritance.* Harvard University Press, Cambridge, Massachusetts.

McClintock B. 1951. Chromosome organization and gene expression. *Cold Spring Harbor Symp. Quant. Biol.* **16:** 13–57.

——— 1984. The significance of responses of genome to challenge. *Science* **226:** 792–800.

McClintock B. and Creighton H.B. 1931. A correlation of cytological and genetical crossing over in *Zea mays. Proc. Natl. Acad. Sci.* **17:** 492–497.

Moore J. 1972a. *Heredity and development,* 2nd ed. Oxford University Press, Oxford.

——— 1972b. *Readings in heredity and development.* Oxford University Press, Oxford.

Morgan T.H. 1910. Sex-linked inheritance in *Drosophila. Science* **32:** 120–122.

Morgan T.H., Sturtevant A.H., Muller H.J., and Bridges C.B. 1915. *The mechanism of Mendelian heredity.* Holt, Rinehart & Winston, New York.

Muller H.J. 1927. Artificial transmutation of the gene. *Science* **46:** 84–87.

Olby R.C. 1966. *Origins of Mendelism.* Constable and Company Ltd., London.

Peters J.A. 1959. *Classic papers in genetics.* Prentice-Hall, Englewood Cliffs, New Jersey.

Rhoades M.M. 1946. Plastid mutations. *Cold Spring Harbor Symp. Quant. Biol.* **11:** 202–207.

Sager R. 1972. *Cytoplasmic genes and organelles.* Academic Press, New York.

Scott-Moncrieff R. 1936. A biochemical survey of some Mendelian factors for flower color. *J. Genetics* **32:** 117–170.

Simpson G.G. 1944. *Tempo and mode in evolution.* Columbia University Press, New York.

Sonneborn T.M. 1950. The cytoplasm in heredity. *Heredity* **4:** 11–36.

Stadler L.J. 1928. Mutations in barley induced by X-rays and radium. *Science* **110:** 543–548.

Sturtevant A.H. 1913. The linear arrangement of six sex-linked factors in *Drosophila* as shown by mode of association. *J. Exp. Zool.* **14:** 39–45.

Sturtevant A.H. and Beadle G.W. 1962. *An introduction to genetics.* Dover, New York.

Sutton W.S. 1903. The chromosome in heredity. *Biol. Bull.* **4:** 231–251.

Wilson E.B. 1925. *The cell in development and heredity,* 3rd ed. Macmillan, New York.

Wright S. 1931. Evolution in Mendelian populations. *Genetics* **16:** 97–159.

——— 1941. The physiology of the gene. *Physiol. Rev.* **21:** 487–527.

第2章 核酸承载遗传信息

本章概要

•

Avery的惊人发现：DNA能够携带遗传特性

•

双螺旋

•

DNA中的遗传信息是由4种核苷酸单元形成的序列所承载的

•

中心法则

•

确定蛋白质合成的方向

•

基因组学时代

早在化学家们注意到存在能够承载遗传信息的特殊分子之前，遗传学家们就已着手探询这种分子了。20世纪30年代，遗传学家就开始推测什么样的分子可能具有基因所需要的稳定性，同时又能有长久的、偶然的突变形式，这是进化的基础。直到40年代中期，似乎还没有直接的方式可以探求基因的化学本质。人们早就知道染色体含有一种特殊的分子，脱氧核糖核酸（DNA），但是没有办法证明这种分子携带遗传信息，相反却认为DNA仅仅是一种分子支架，支撑某种尚未发现的、专门携带遗传信息的蛋白质。一般认为基因是由氨基酸组成的，因为当时，氨基酸看起来是唯一具有足够复杂而能携带遗传信息的生物分子。

因此，探询基因在细胞中的作用，进而了解基因的本质就颇有意义了。在20世纪40年代早期，以George W. Beadle和Edward Tatum为先锋对脉孢菌（*Neurospora*）的研究中，发现越来越强的证据支持30年前Garrod提出的假说，即基因的任务就是控制特定酶的合成（一个基因一个酶假说）。考虑到当时已经知道所有已知的酶都是蛋白质，关键的问题是基因参与蛋白质合成的方式。从种种推测开始，最简单的假说就是基因中的遗传信息决定蛋白质多肽链中20种不同氨基酸的顺序。

验证这一假说，直觉几乎没有什么用，即使对最好的生物化学家也如此，因为用酶作为工具检测添加到多肽链中的氨基酸的顺序是不合逻辑的。如果那样的话就应该需要同蛋白质所含氨基酸数目相同的排序酶，才能合成蛋白质。但是，当时所有已知的酶自身也都是蛋白质（现在我们知道RNA在个别情况下也可以发挥酶的功能），那么（如果按上述逻辑），就需要另外的排序酶来合成排序酶。显然这是一个自相矛盾的怪圈，除非我们假定有一种异乎寻常地互相衔接的合成，其中某个特定的蛋白质具有许多不同的酶的特性。在这样的假设前提下，我们才有可能看到（将是非常困难的）一个可以行

使功能的细胞。然而，大部分蛋白质不可能执行多种功能。事实上，现在所有的知识指向相反的方向，即一种蛋白质一种功能。

Avery 的惊人发现：DNA 能够携带遗传特性

DNA 有可能是主要的遗传分子的证据，从肺炎致病菌的研究中出人意料地显示出来。1928 年，英国微生物学家 Frederick Griffith 做出了开创性的观察：当非致病肺炎链球菌菌株与热灭活的同种致病菌混合后，成为致病的细菌。这种从非致病菌向致病菌的转化（transformation）表明了遗传的改变，这一现象通过用新成为致病菌的后代菌株转化其他的非致病细菌可以显现出来。这就产生了一种可能性推测，即当病原细胞被热灭活时，其遗传组分并没有被破坏。而且，一旦它们（遗传组分）从热灭活的细胞中释放出来，还能穿过活的受体细胞的细胞壁并与受体的遗传组分进行重组（图 2-1）。随后的研究验证了这一遗传的解释。致病性显示了荚膜基因的功能，该基因编码一个含有糖类的荚膜的合成中的关键酶，在大多数致病肺炎菌上都有这种荚膜。当 *S*（smooth，平滑的）等位基因存在时，致病性必需的荚膜就会在细胞的外围形成（形成荚膜的同时也会使细菌所形成的菌落呈现平滑的外观）；当 *R*（rough，粗糙的）等位基因存在时，没有荚膜形成，相应的细胞是非致病的，这些细胞形成的菌落的边缘是平滑的。

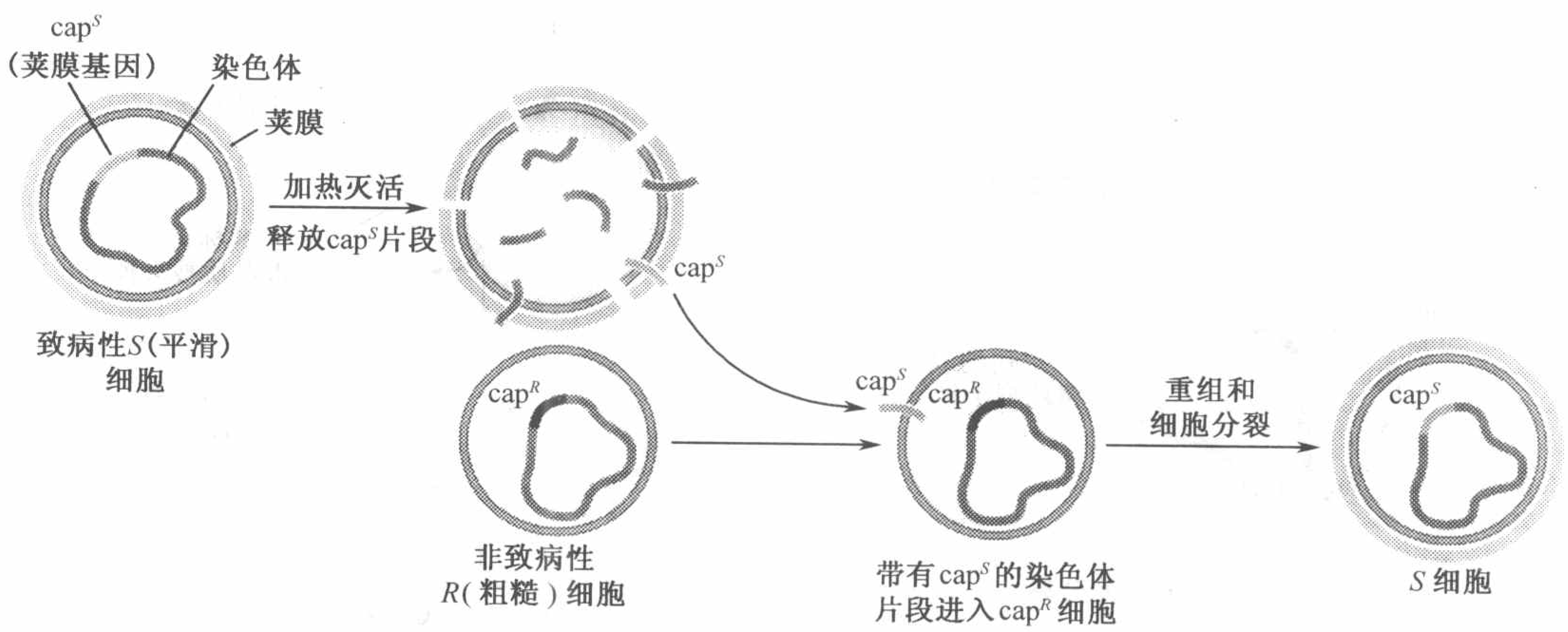


图 2-1　通过遗传上不同的热灭活菌株，一种细菌（肺炎双球菌）的遗传特性得到转化。本图表明 *R* 型菌株接受来自热灭活的 *S* 型菌株的包含荚膜基因的染色体片段。由于大多数 *R* 型细胞接受的是别的染色体片段，因此特定基因的转化率通常低于 1%。

在 Griffith 观察到上述现象的最初几年内，又发现灭活后的细菌抽提物也具有诱发遗传转化的能力，他们立即展开了确定这种转化因子的化学本质的工作。当时，大部分生化学家仍然相信基因应该是蛋白质。所以，在历经十几年的研究后，当美国微生物学家 Oswald T. Avery 和他在纽约的洛克菲勒研究所的同事 Colin M. MacLeod 和 Maclyn McCarty 在 1944 年宣布他们的重要发现，即有活性的遗传物质是 DNA 时（图 2-2），所有人都大吃一惊。支持他们结论的关键实验是：所获得的高度纯化的活性组分的转化

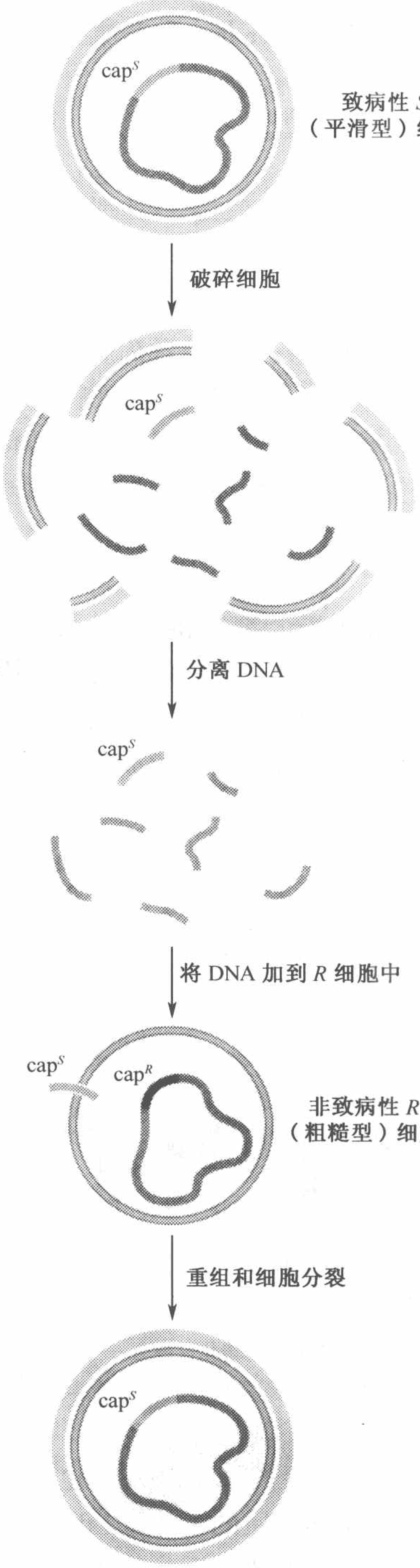


图 2-2 分离纯化转化因子。（来源：Stahl F. W. 1964. The mechanics of inheritance，Fig. 2.3 Copyrigh © 1964. Reprinted by permission of Pearson Education，Inc.，Upper Saddle River. NJ.）

活性可以被来自胰腺的脱氧核糖核酸酶所破坏。该酶是刚刚纯化的，可以特异性地把 DNA 分子降解成核苷酸，但是对蛋白质或 RNA 的完整性没有任何影响。添加胰腺的核糖核酸酶（可以降解 RNA）或各种蛋白酶（破坏蛋白质）对转化活性都没有影响。

病毒基因也是核酸

对病毒和被病毒感染的细胞的化学研究为上述理论提供了更重要的支持证据。到

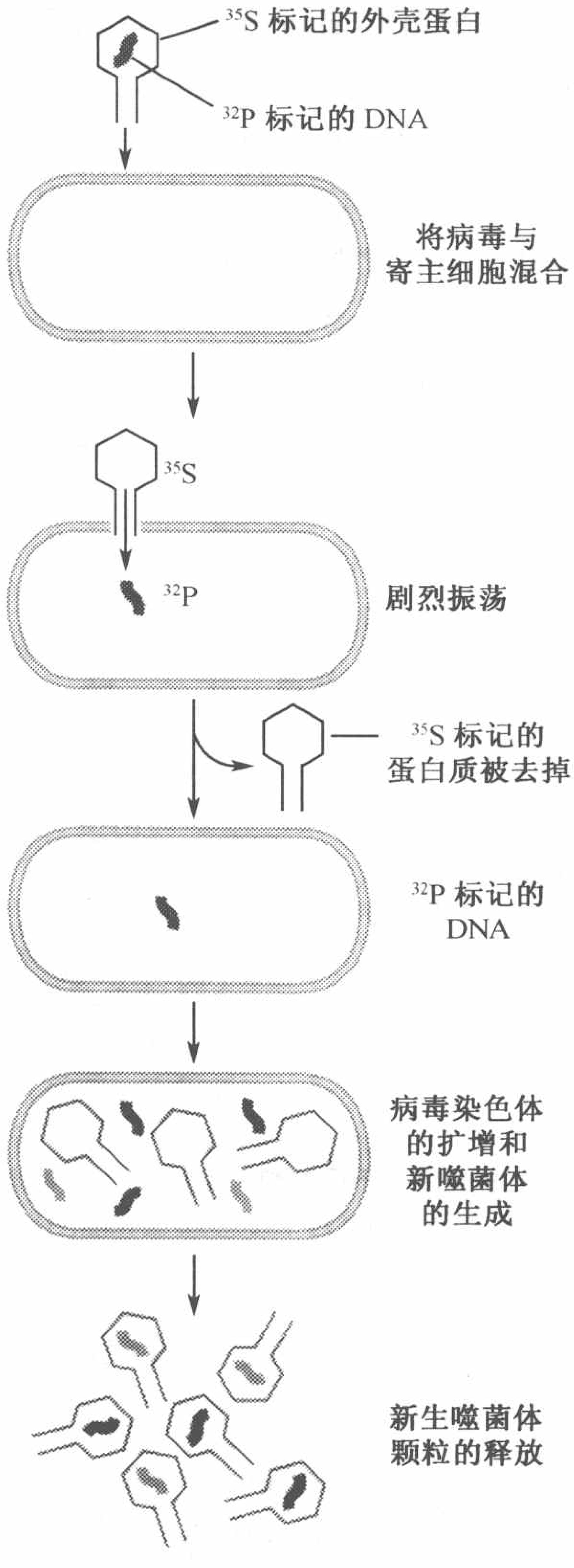


图 2-3 图示只有 T2 噬菌体的 DNA 组分携带遗传信息，蛋白质外壳只是保护性。

1950年，已经可以获得若干相当纯的病毒，并能够检测病毒中含有什么类型的分子。这项工作获得了一个非常重要的且具有广泛意义的结论，即所有的病毒都含有核酸。因为在当时已经越来越意识到病毒是含有遗传物质的，所以立即就提出了一个问题，是不是核酸就是病毒基因的载体呢？为了回答这一问题，1952年，Alfred D. Hershey 和 Martha Chase 在纽约长岛的冷泉港实验室完成了一项实验，用同位素研究细菌病毒 T2［一种噬菌体（bacteriophage，phage）］的增殖，这种噬菌体由 DNA 核心区和由多种不同的蛋白质分子集合形成的、起保护作用的外壳组成。他们将蛋白质外壳用放射性同位素^{35}S标记，核心区 DNA 用放射性同位素^{32}P标记。然后用被标记的病毒跟踪蛋白质和核酸在噬菌体增殖过程中的走向，主要是检测母本病毒标记的哪种原子进入寄主细胞并出现在后代噬菌体中。

从这些实验获得的结果清楚无误地表明：在后代噬菌体中能够检测到大量的母本噬菌体的核酸，而没有任何母本蛋白质（图 2-3）。此外，还证明了几乎没有母本蛋白质进入细菌，而只能吸附在细菌的外壳上，当 DNA 进入细菌细胞后这些蛋白质就没有任何功能了。这一点由以下实验巧妙地证明：在 DNA 进入细胞后剧烈振荡侵染的细菌，结果吸附在细菌上的外壳蛋白质脱落但并不影响细菌形成新的噬菌体颗粒。

利用某些病毒，现在可以进行更令人信服的实验。例如，从小鼠多瘤病毒纯化的 DNA 可以进入小鼠细胞，启动病毒繁殖的周期并产生数以千计的新的多瘤病毒颗粒。所以，病毒蛋白质的基本功能就是在细胞间移动过程中保护其遗传成分——核酸。

双螺旋

在用 X 射线分析蛋白质结构的同时，有一少部分科学家开始尝试分析 DNA 的 X 射线衍射图谱。1938 年 William Astbury 用 Ola Hammarsten 和 Torbjorn Caspersson 提供的 DNA 样品拍摄到了第一张衍射图谱。但直到 20 世纪 50 年代初期，Maurice Wilkins 和 Rosalind Franklin 才获得了高质量的 X 射线衍射照片（图 2-4）。从这些照片来看，DNA 的结构应该是螺旋形的，而且不止一条多核苷酸链，不是 2 条就是 3 条。与此同时，将 DNA 分子连接起来的共价键也清楚了。1952 年，一群在 Alexander Todd 实验室工作的有机化学家证明：将 DNA 的核苷酸结合在一起的是 3′-5′的磷酸二酯键（图 2-5）。

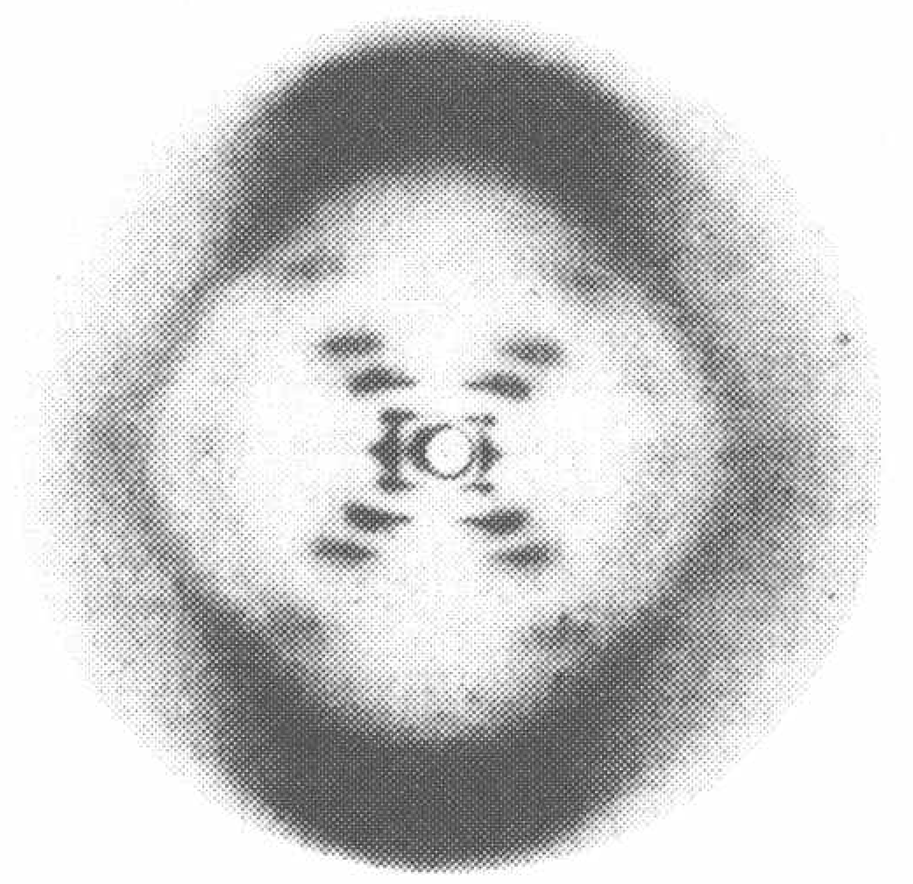


图 2-4 用来阐明 DNA 结构的关键的 X 射线照片。这张照片由 Rosalind Franklin 于1952～1953 年冬天在伦敦 King 学院拍摄，它证明了 DNA 是螺旋形结构的猜想。照片中心的十字形的 X 射线反射图形表明了螺旋的形式（X 射线照片上暗的部位）；位于顶端和底部的很深的黑色区域表明相距 3.4Å 的嘌呤和嘧啶碱基是规则的相邻叠加的，并垂直于螺旋轴。（来源：Franklin R. E. and Gosling R. 1953. *Nature* 171：740，经许可）

出于对 Linus Pauling 的 α 螺旋蛋白基序（motif）(将于本书第 5 章涉及）感兴趣，William Cochran、Francis H. Crick 和 Vladimir Vand 在 1951 年建立了一种很好的关于螺旋分子的衍射理论，使通

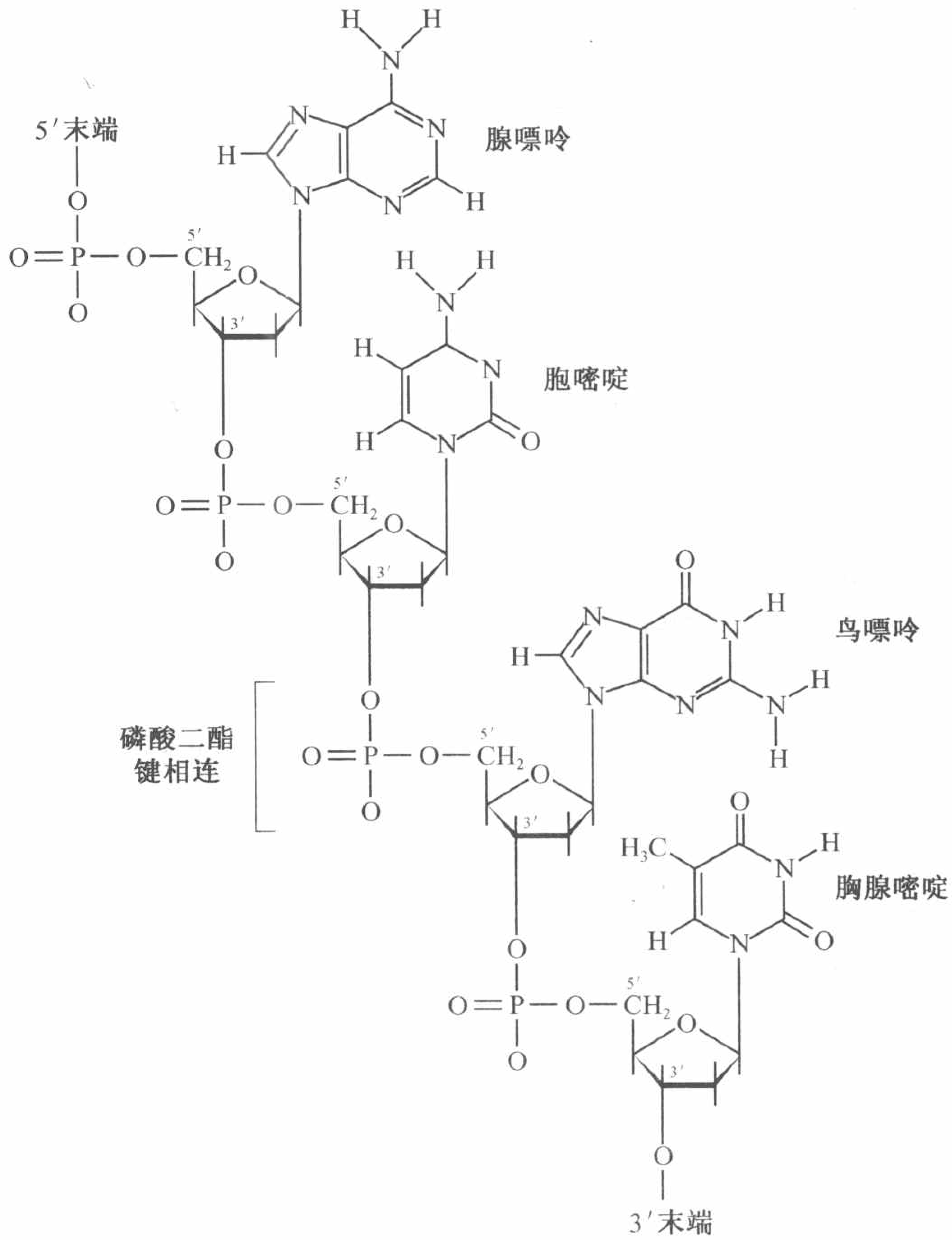


图 2-5 DNA 多聚核苷酸的一部分，图示连接核苷酸的 3′-5′的磷酸二酯键的连接方式。磷酸基团将一个核苷酸的 3′碳原子与另一个核苷酸的 5′碳原子相连。

过实验探索可能的 DNA 结构更简单。1953 年，克里克和沃森提出了正确的互补双螺旋（见第 6 章）理论，当时他们是在 Max Perutz 和 John Kendrew 的实验室工作。正确答案的获得主要依赖于发现了与 Wolkins 和 Franklin 的 X 射线衍射数据相符合并且从立体化学角度来讲最可能的构象。

在双螺旋中，两条反向的 DNA 链由碱基对之间的氢键（一种非共价的化学弱键，见第 3 章）结合在一起（图 2-6）。这种碱基配对是高度特异的：腺嘌呤只与胸腺嘧啶配对，而鸟嘌呤只与胞嘧啶配对。在双螺旋 DNA 分子中，A 残基的数目一定与 T 残基的数目相等，同样地，G 的数目一定与 C 的数目相等（见框 2-1）。结果是，双螺旋的两条链的碱基序列有互补的关系，其中任何一条 DNA 链的序列都严格地决定了其对应链的序列。

双螺旋结构的发现对许多遗传学家分析数据的方式带来了意义深远的革命。基因不再是一种神秘的东西，它的活动只能用遗传实验来研究。相反，它已经变成了一个化学家可以客观面对的真实的分子物体，就像其他的小分子，如丙酮酸盐和 ATP 一样。然而，最令人振奋的还不仅仅是破译了结构，而且还来自结构的内涵。在知道答案之前，人们一直在担心（基因的结构的）答案难以揭示基因复制和功能；值得庆幸的是，这一

答案非常令人振奋。两条互补的相互缠绕的链本身就意味着一条链能够为另一条链的合成提供一个特定的表面（模板）（图 2-6）。如果这一假说成立，那么困扰遗传学家们多年的关于基因复制的基本问题，就从理论上得到解决了。



图 2-6 DNA 的复制。新合成的链用橙色表示。

框 2-1 Chargaff 定律

生化学家 Erwin Chargaff 用纸层析技术分析 DNA 的核苷酸组成。1949 年，他提出的数据表明不但 4 种不同的核苷酸的含量不同，而且 4 种核苷酸的比率在不同的物种中也不一样（框 2-1 表 1）。这些发现意味着一种可能性，即 DNA 分子中核苷酸的精确排列承载着其遗传特性。

Chargaff 实验同时也表明，4 种碱基的相对比率不是随机的。所有 DNA 样品中，A 残基的数目与 T 残基的数目相等，而 G 残基的数目等于 C 的数目。另外，不论 DNA 的来源如何，嘌呤与嘧啶的比值都大体上等于 1（嘌呤=嘧啶）。在知道 DNA 的三维结构之前是难以理解 A=T 和 G=C 的关系（Chargaff 定律）的重要意义的。

框 2-1 表 1 莫定 Chargaff 定律的数据

来源	A∶G	T∶C	A∶T	G∶C	嘌呤∶嘧啶
牛	1.29	1.43	1.04	1.00	1.1
人	1.56	1.75	1.00	1.00	1.0
母鸡	1.45	1.29	1.06	0.91	0.99
鲱鱼	1.43	1.43	1.02	1.02	1.02
小麦	1.22	1.18	1.00	0.97	0.99
酵母	1.67	1.92	1.03	1.20	1.0
嗜血流杆菌	1.74	1.54	1.07	0.91	1.0
大肠杆菌 K2 菌株	1.05	0.95	1.09	0.99	1.0
禽类结核杆菌	0.4	0.4	1.09	1.08	1.1
沙雷氏菌	0.7	0.7	0.95	0.86	0.9
杆菌	0.7	0.6	1.12	0.89	1.0

来源：After Chargaff E. et al. 1949. *J. Biol. Chem.* 177：405.

合成 DNA 的聚合酶的发现

为严格地证明一条 DNA 链是指导其互补的 DNA 链合成的模板，还需要建立试管

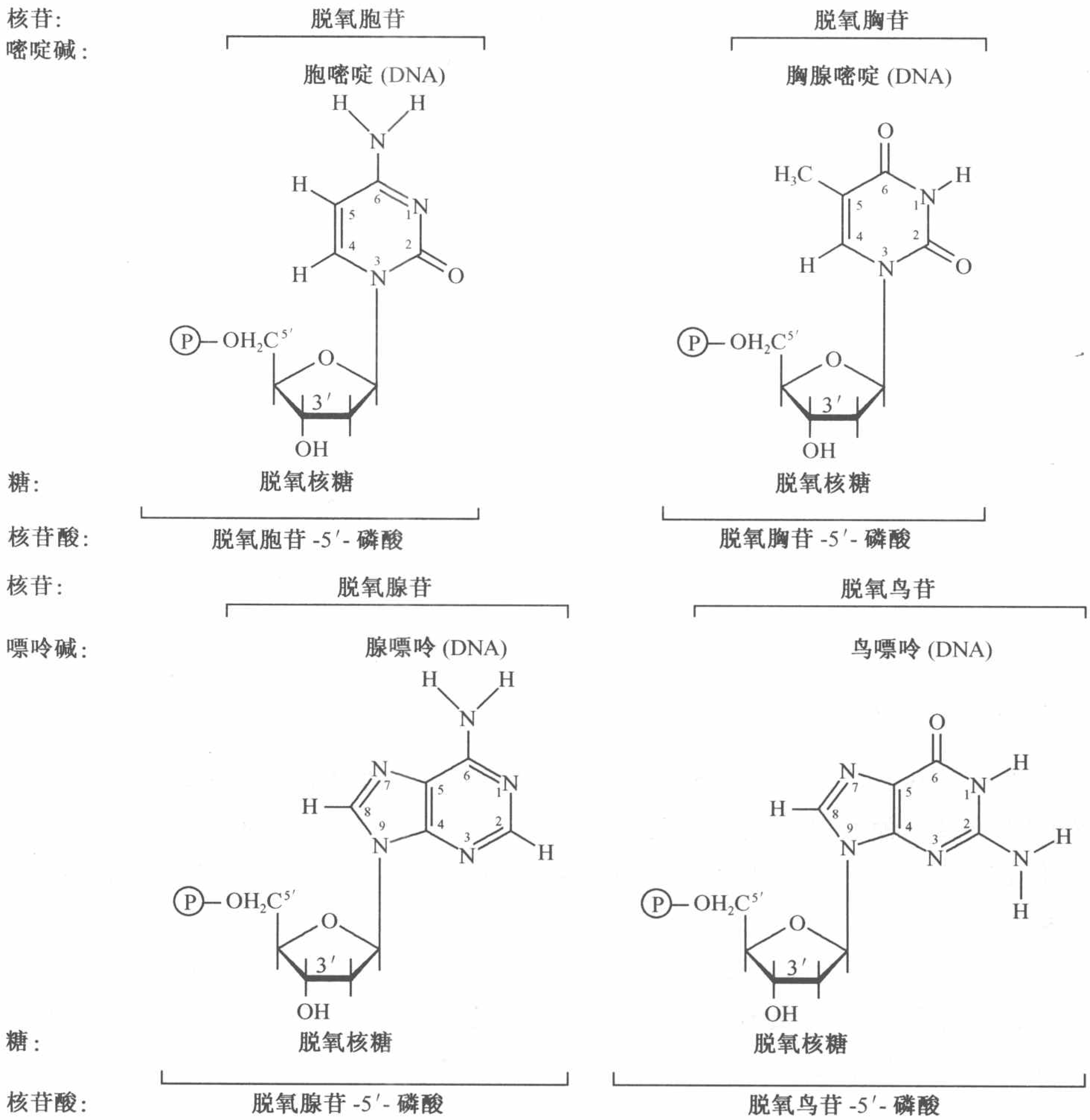


图 2-7 DNA 的核苷酸。图中显示的是 4 种核苷酸的不同结构。

中（离体）DNA合成系统，而实际上，这一体系在很短的时间内被建立，出乎分子遗传学家们的预料。当时，分子遗传学家与生物化学家的工作相距很远，后者相当擅长酶纯化方面的工作。利用生化技术破解DNA复制之谜的领军人物是美国生化学家Arthur Kornberg，他在1956年用无细胞的细菌抽提物证明了DNA的合成。在随后的几年中Kornberg又进一步证明了催化完成DNA组成元件的连接需要一种特定的聚合酶。Kornberg的研究揭示了组成DNA的核苷酸元件是富含能量的前体物质（dATP、dGTP、dCTP和dTTP，图2-7）。进一步的实验鉴定了一种单链的多肽——DNA聚合酶Ⅰ（DNA Pol Ⅰ），它能够催化新的DNA链的合成。DNA聚合酶Ⅰ将核苷酸前体用3′-5′的磷酸二酯键连接起来（图2-8）。这种反应需要DNA的存在，其作用是将4种核苷酸按顺序排列成最终的多聚核苷酸产物。

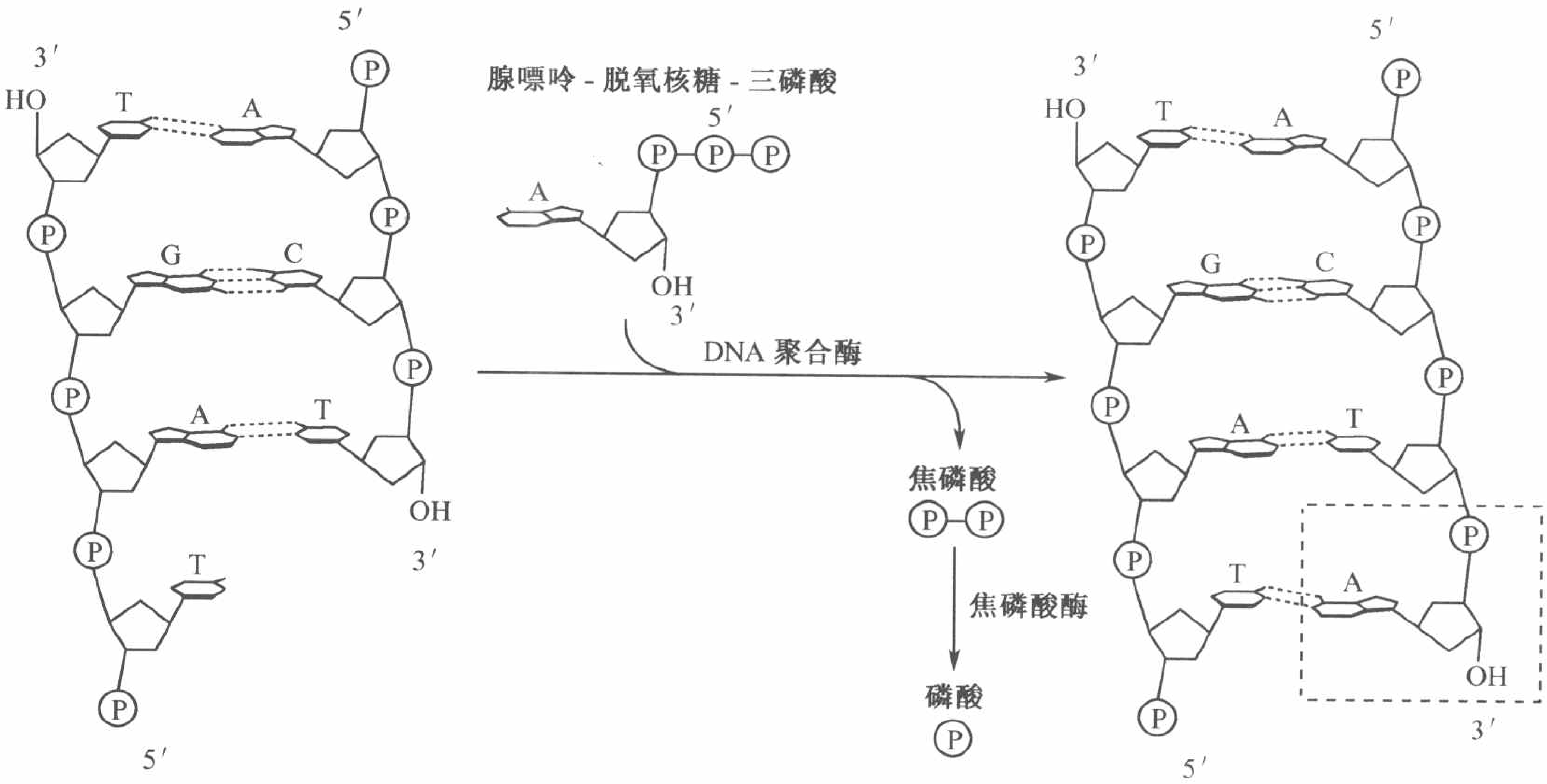


图2-8 由DNA聚合酶Ⅰ催化的DNA链的酶促合成反应。图示将核苷酸添加到不断增长的DNA链上去的反应是由DNA聚合酶催化的。尽管DNA聚合酶本身就能催化细胞中的DNA合成，释放的焦磷酸是由焦磷酸酶转化为二磷酸的，更加有利于核苷酸添加的正向反应。

DNA聚合酶Ⅰ依赖DNA模板来决定所合成的DNA的顺序。这一点首先由以下实验证明：用含不同数量的A∶T和G∶C碱基对的DNA分子作模板，进行酶促合成。在所有的实验中，酶促合成的产物都与模板DNA的碱基比率相同（表2-1）。在无细胞体系的合成中，没有蛋白质或其他任何形式的分子的形成，完全排除了其他非DNA的组分作为遗传特性的中间载体的可能性。所以，毫无疑问DNA就是合成自身的直接模板。

支持DNA复制过程中双链分开的实验证据

1958年，在Kornberg研究的同时，当时在加州理工学院的Matthew Meselson和Frank W. Stahl进行了一个非常聪明的实验：他们将子代DNA分子分离，证明了双螺旋的两条链在DNA复制过程中一直是分开的（图2-9）。实验的成功部分归因于用^{15}N

表 2-1　酶促合成的 DNA 产物及 DNA 模板中的碱基组成比较

DNA 模板来源	酶促合成产物的碱基组成				A+T/G+C	A+T/G+C
	腺嘌呤	胸腺嘧啶	鸟嘌呤	胞嘧啶	产物中	模板中
球菌（细菌）	0.15	0.15	0.35	0.35	0.41	0.39
产气杆菌（细菌）	0.22	0.22	0.28	0.28	0.80	0.82
大肠杆菌	0.25	0.25	0.25	0.25	1.00	0.97
牛胸腺	0.29	0.28	0.21	0.22	1.32	1.35
噬菌体 T2	0.32	0.32	0.18	0.18	1.78	1.84

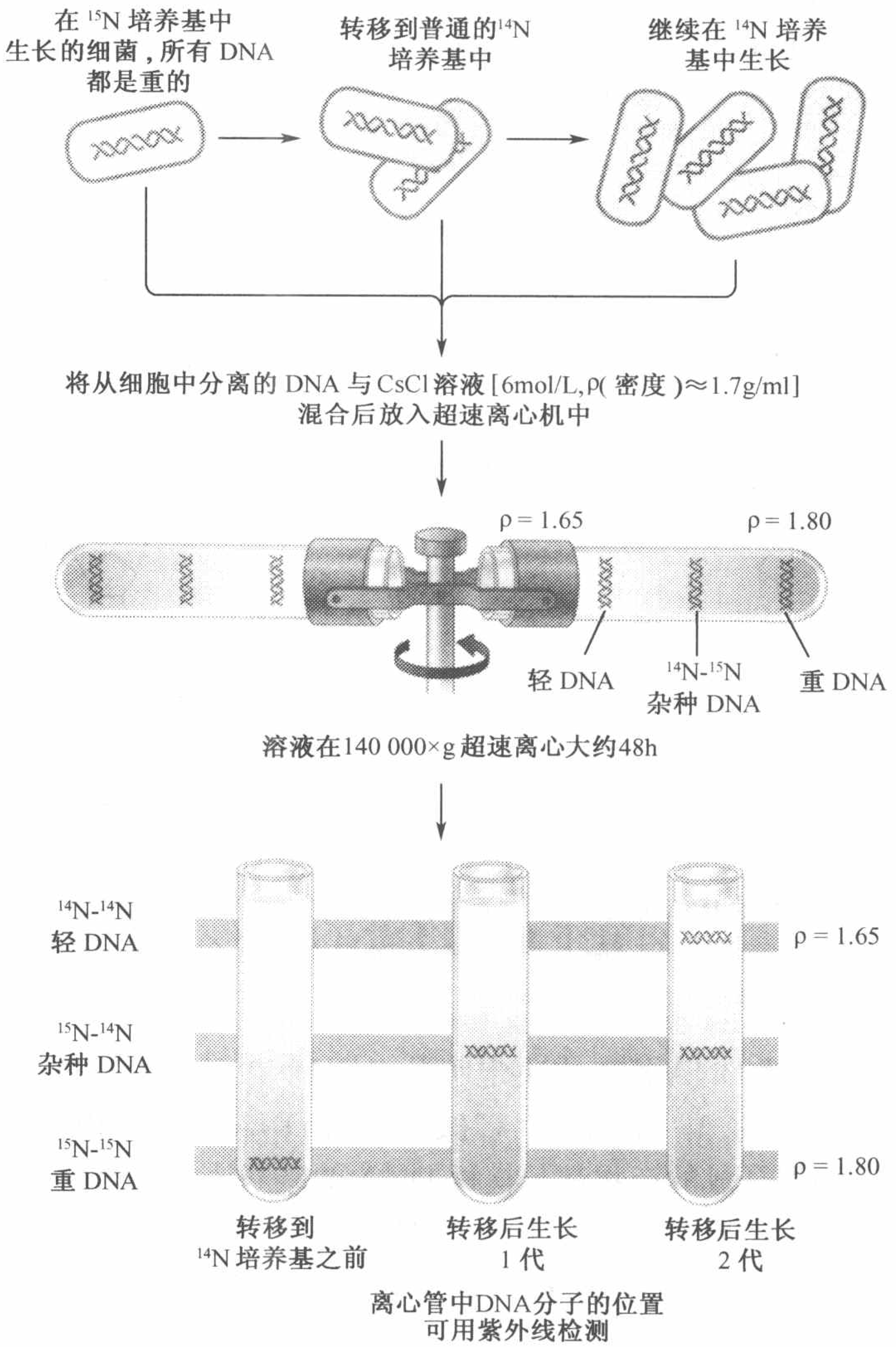


图 2-9　用氯化铯密度梯度证明 DNA 复制过程中互补链的解开。

重同位素作为标签来区分亲代和子代 DNA 链。生长在含重同位素^{15}N的培养基中的细菌的DNA 比在正常的^{14}N培养基中的细菌的 DNA 要重一些。另外，实验的成功还部分地归因于用重盐（如氯化铯）进行密度梯度离心分离轻、重 DNA 实验技术的建立。当施加高离

心力时，在离心管底部的溶液密度会变得更高（旋转时，它们是离旋转轴最远的位置）。如果选择的初始溶液密度合适，DNA 分子就会移动到离心管的中部，停留在与盐溶液的密度相同的区域。此时，同两条链都完全是^{14}N前体的“轻轻 DNA”（LL DNA）相比，两条链都完全是^{15}N的“重重 DNA”（HH DNA）会在较高密度区形成一条带（离离心管底部更近）。当把含有重 DNA 的细菌转移到普通培养基（含^{14}N）上培养时，用来进行 DNA 合成的核苷酸前体是轻的，所以转移后新合成的 DNA 同转移前的 DNA 是可以区分的。

如果 DNA 的复制涉及解链过程，在普通培养基上生长的携带重 DNA 的细菌在不同阶段后的 DNA 分子的密度就可以推测出来。一个世代后，所有的 DNA 分子应该含有一条重链和一条轻链，也就是居中的杂种密度（“重轻 DNA”，HL DNA）。这正是 Meselson 和 Stahl 所观察到的现象。同样地，生长两代以后，一半的 DNA 分子是轻的，另一半是杂种，与解链的预测完全符合。一定要注意：DNA 在从细菌里纯化出来时，会被断裂成小片段。这就保证了大多数 DNA 不是被完全复制，要么不被复制。如果细菌基因组保持完整，很可能会有很多部分复制的中间密度的分子（既不是 HH、HL，也不是 LL）存在。

由此 Meselson 和 Stahl 的实验证明了 DNA 复制是半保留的过程，在复制过程中双螺旋中的一条链保持不变（也就是“保留”），两条亲本链分别分配到两个子代分子中（半保留复制中的“半”）。这些实验排除了当时的另外两种模型：保留和分散复制模型（图 2-10）。保留模型认为，亲本的两条链不分开，而新形成的两条链则形成全新的 DNA 分子。按照这一模型，细胞生长一代后就应该有轻 DNA 形成。在当时许多人都

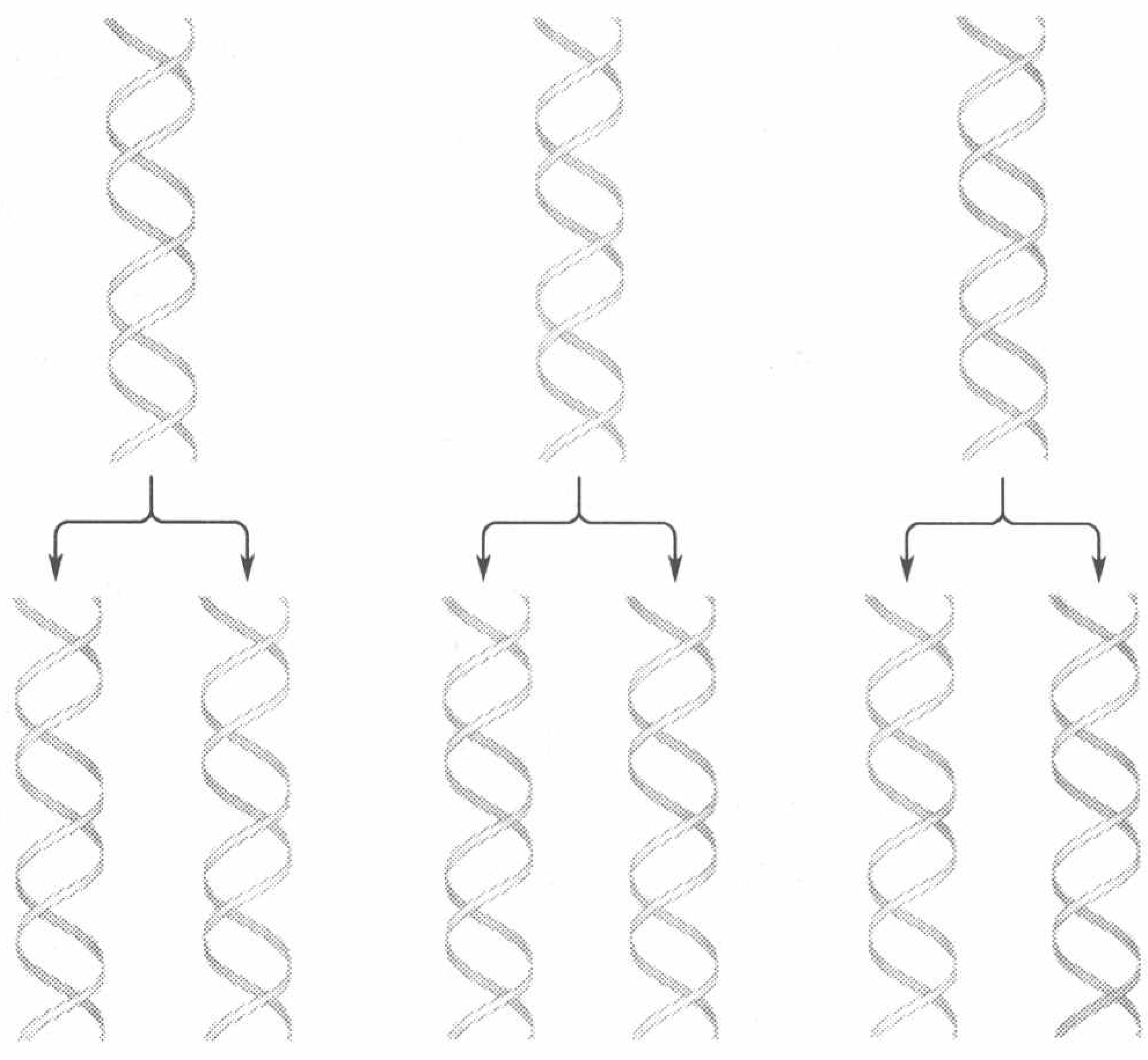


图 2-10 3 种可能的 DNA 复制机制。了解了 DNA 结构之后，提出了几种解释 DNA 复制的模型，本图介绍了 3 种。由 Meselson 和 Stahl 提出的实验与其他模型不同，清楚地证明了 DNA 是半保留复制的。

特别倾向的分散复制模型中，DNA 链首先被打成碎片，片段的长度可能只有 10 个碱基对，利用这些片段合成同样短的 DNA 片段。然后这些短片段再连接起来形成完整的 DNA 链。按照这个复杂的模型，DNA 链将是由新老 DNA 共同组成的（即非保留的），只能在多代生长之后才会有轻 DNA 出现。

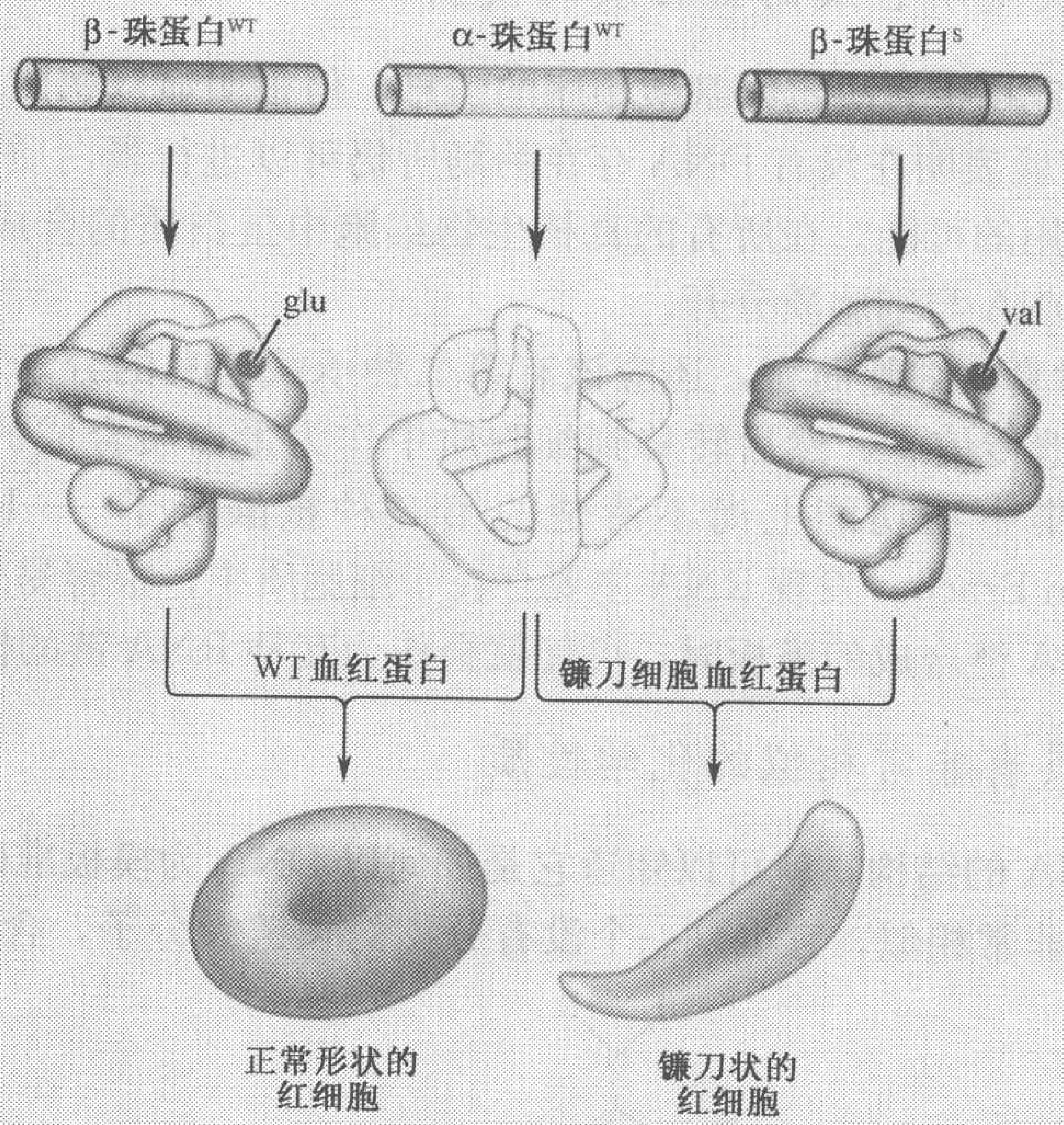
DNA 中的遗传信息是由 4 种核苷酸单元形成的序列所承载的

双螺旋的发现有效地终止了 DNA 是否是主要遗传物质的争论。在用实验证明 DNA 的复制涉及解链之前，分子遗传学家就将研究重点转向了解 DNA 中的遗传信息是如何决定合成蛋白质的氨基酸顺序的（见框 2-2）。因为所有的 DNA 链都能够形成双螺旋，其遗传特性的核心肯定是隐含在由 4 种核苷酸单元形成的线性顺序中。所以，作为携带遗传信息的实体，当时人们就正确地认为 DNA 分子应该是由 4 个字母（A、G、C 和 T）构成的很长的单词（在后面的章节中我们还将看到，现在更应该说 DNA 是很长的句子）。即使只有 4 个字母，可能的 DNA 顺序的数目（4^N，其中 N 是序列中字母的个数）即使对最小的 DNA 分子来说，也是非常的庞大，实际上可能存在的遗传信息是无限的。现在我们知道一个典型的细菌的基因由大约 1000 个碱基对组成的，这样大小的一段 DNA 可能产生的基因数目是 4^{1000}，这个数目比所有生物中已知的基因数目都大几个数量级。

框 2-2　基因控制蛋白质中氨基酸顺序的证据

基因（DNA）控制氨基酸顺序的第一个实验证据来自于对患有镰状细胞贫血症遗传病的患者的血红蛋白的研究。如果一个人的 2 条同源染色体都带有 β 珠蛋白基因（编码组成血红蛋白的两条肽链中的一个）的 S 型等位基因，就会导致严重的贫血症，其典型特征是患者的红细胞呈镰刀形。如果 β 珠蛋白基因的一个等位基因是 S 型，贫血症状就相对较轻，红细胞也接近正常的形状。红细胞中血红蛋白的类型似乎与遗传特征相符合。在 SS 的情况下，血红蛋白不正常，特征是其溶解性不同，而在＋S（一个等位基因是野生型——译者注）的情形下，一半血红蛋白是正常的，另一半是镰刀形。

野生型血红蛋白分子是由两种多肽链组成的，α 链和 β 链（框 2-2 图 1）。每一条肽链的分子质量大约是 16 100Da。每一个血红蛋白分子中有两条 α 链和两条 β 链，总的分子质量大约为 64 400Da。α 链和 β 链是由不同的基因控制的，所以一个突变只能影响一条链，α 链或者 β 链，但不会同时影响二者。1957 年，剑桥大学的 Vernon M. Ingram 证明了镰状血红蛋白与正常血红蛋白的区别在于 β 链的一个氨基酸的改变：第 6 个氨基酸位置，由正常的谷氨酸变成了缬氨酸。除此之外，正常和突变血红蛋白的所有氨基酸序列都相同，因为这种氨基酸顺序的改变只在带有 S 型的 β 珠蛋白基因的患者中才会出现。一个最简单的假设是 S 等位基因决定了 β 珠蛋白基因的突变。随后对其他类型贫血症 β 珠蛋白氨基酸序列的分析完全支持这种解释；序列分析表明，每一种特定的贫血症都有多肽链特定位置上一个氨基酸的改变（框 2-2 图 2）。



框 2-2 图 1　野生型和镰状细胞血红蛋白的形成（血红蛋白结构的来源：图自 Irving Gels。版权为 Howard Hughes Medical Institute 所有，未经许可不得复制）。

α链								
位置	1	2	16	30	57	58	68	141
氨基酸	Val	Leu	Lys^+	Glu^-	Gly	His^+	AspN	Arg
血红蛋白的突变体								
Hb I			Asp^-					
Hb G Honolulu				GluN				
Hb Norfolk					Asp^-			
Hb M Boston						Tyr		
Hb G Philadelphia							Lys^+	

β链										
位置	1	2	3	6	7	26	63	67	125	150
氨基酸	Val	His^+	Leu	Glu^-	Glu^-	Glu^-	His^+	Val	Glu	His^+
血红蛋白的突变体										
Hb S				Val						
Hb C				Lys^+						
Hb G San José					Gly					
Hb E						Lys^+				
Hb M Saskatoon							Tyr			
Hb Zürich							Arg^+			
Hb M Milwaukee-1								Glu^-		
Hb D β Punjab									GluN	

框 2-2 图 2　一些已经确认的人类血红蛋白突变体的氨基酸替换。

DNA 不是直接决定所合成的蛋白质的模板

虽然 DNA 一定携带着决定氨基酸顺序的信息，但很明显双螺旋本身不可能是蛋白质合成的模板。实验表明在没有 DNA 存在的场所仍可以进行蛋白质的合成，所以排除了 DNA 起直接作用的可能。在所有的真核生物细胞中蛋白质的合成发生在细胞质中，而核膜将染色体 DNA 与细胞质分开。

所以，至少对真核细胞而言，还应该有第二种承载信息的分子存在，由它负责从 DNA 上获得特定遗传信息，然后转移到细胞质中作为蛋白质合成的模板。从一开始，人们的注意力就集中在功能上尚不清楚的第 2 种核酸分子——RNA 上。Torbjorn Caspersson 和 Jean Brachet 发现 RNA 主要存在于细胞质中；很容易想像，当一条 DNA 链不用于互补 DNA 链合成的模板时，可能就会作为互补 RNA 链的模板。

RNA 与 DNA 具有非常相似的化学性质

通过了解 RNA 的结构，就可以知道它是如何以 DNA 为模板准确合成的。RNA 与 DNA 的化学性质非常相似，它也是一个没有分支的长链大分子，含有 4 种核苷酸，以

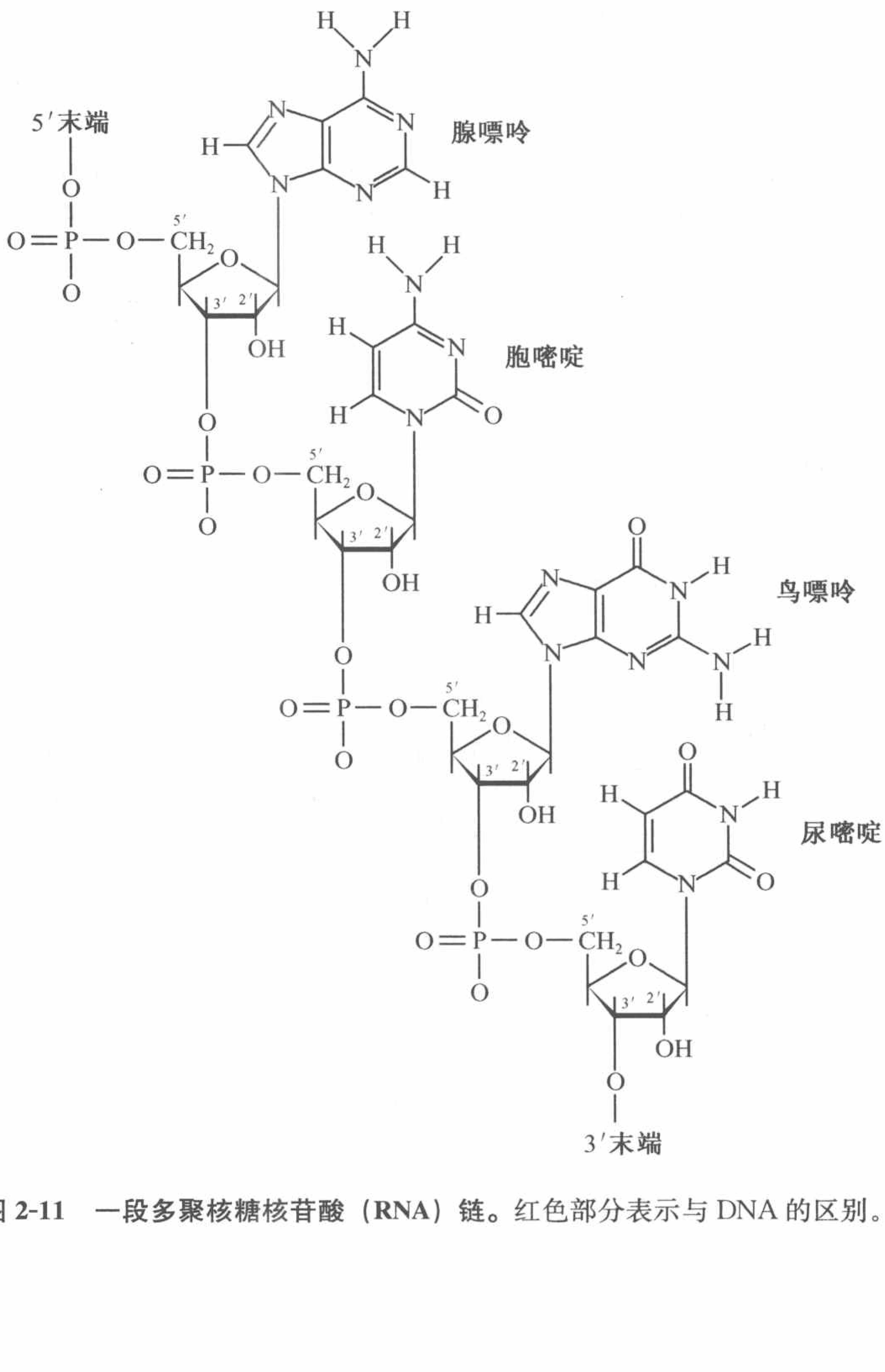


图 2-11　一段多聚核糖核苷酸（RNA）链。红色部分表示与 DNA 的区别。

3′-5′磷酸二酯键相连（图 2-11）。在化学组成上，RNA 与 DNA 有两个差异，一个是糖组分的微小差异（图 2-12）：DNA 中为脱氧核糖，而 RNA 中为核糖，除多一个 OH（羟基）外，核糖与脱氧核糖相同；第二个差异是 RNA 分子没有胸腺嘧啶，而含有结构相似的尿嘧啶。尽管存在这些差异，多聚核糖核苷酸也有形成 DNA 那种互补双螺旋的能力。额外的羟基以及尿嘧啶中缺少胸腺嘧啶中才有的甲基都不影响 RNA 通过碱基配对形成双螺旋结构的能力。然而，与 DNA 不同，在细胞中 RNA 主要以单链分子存在。如果有双链 RNA 螺旋形成，大多数也是由同一单链 RNA 分子的两个部分形成的。

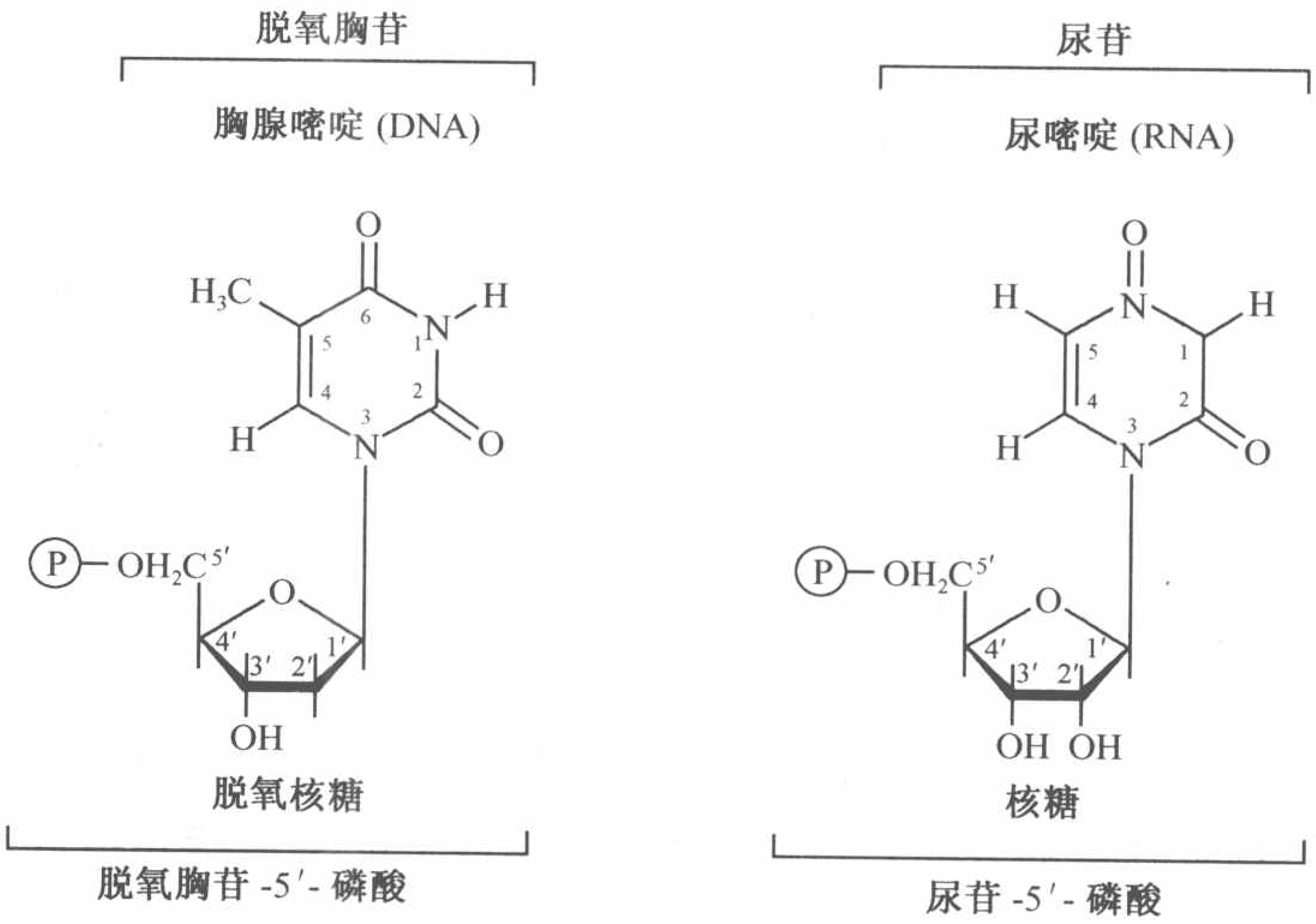
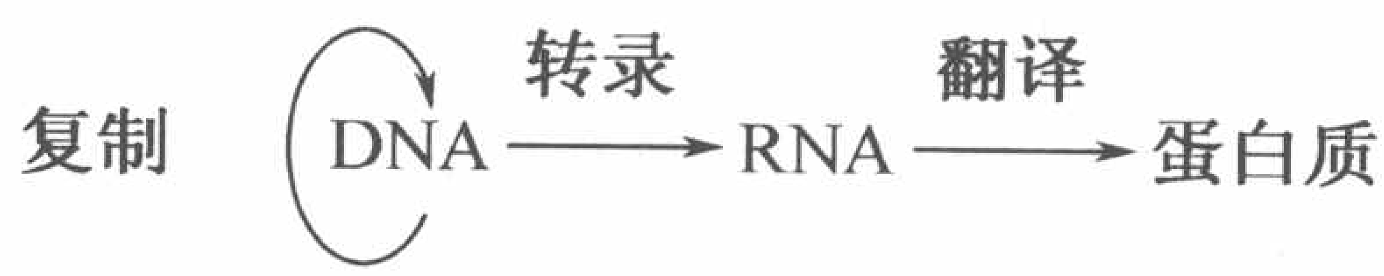


图 2-12 RNA 与 DNA 核苷酸之间的区别。左图为 DNA 的核苷酸，旁边为 RNA 的核苷酸。所有 RNA 的核苷酸含有核糖（而 DNA 中为脱氧核糖），核糖的 C2 位有一个羟基（用红色表示）。此外，在碱基组成上，RNA 中为尿嘧啶，而 DNA 为胸腺嘧啶，尿嘧啶在嘧啶环的 5 位上有羟基（用红色表示），而胸腺嘧啶在该位置是甲基。其余 3 个碱基相同。

中心法则

到 1953 年秋，人们已经普遍接受了如下的工作假说，那就是染色体 DNA 是 RNA 分子的模板，合成后的 RNA 分子转运到细胞质中，在那里决定蛋白质中氨基酸的顺序。1956 年，Crick 提出将遗传信息的传递途径称为中心法则（central dogma）。



其中，箭头表示遗传信息的传递方向。环绕 DNA 的箭头代表 DNA 是自我复制的模板。DNA 与 RNA 之间的箭头表示 RNA 的合成（转录，transcription）是以 DNA 为模板的。相应地，蛋白质的合成（翻译，translation）是以 RNA 为模板的。更为重要的是，

最后两个箭头表示的过程，只能单方向存在，也就是说，不能由蛋白质模板合成 RNA。也难以想像由 RNA 模板合成 DNA。蛋白质不能作为 RNA 的模板已经经受了时间的验证，然而，正如我们将在第 11 章中要提到的，有时 RNA 链确实可以作为互补的 DNA 的模板。这种与正常信息流的方向相反的事例，与大量的以 DNA 为模板合成的 RNA 分子相比，是非常少见的。所以，大约 50 年前提出的中心法则在今天看来仍然是基本正确的。

Crick 的适配子假设

最初，人们简单地认为用于合成蛋白质的 RNA 模板，可以折叠并在其表面形成适合于 20 种不同氨基酸的空腔（cavity）。这些空腔应该有一定的形状，只适合于一种氨基酸，在蛋白质合成过程中，RNA 就是以这种方式决定氨基酸排列的顺序。然而到了 1955 年，Crick 对这种传统的想当然的观点产生了怀疑，他认为该模型预测的结果不可能出现。这是因为，第一，RNA 的 4 种碱基（A、U、G、C）上的特异化学基团应该主要与水溶性基团相互作用，而许多氨基酸（如 Leu、Val、Phe）的侧链基团却强烈地倾向于与非水溶性基团（疏水基团）发生作用。第二，即使 RNA 分子能够折叠并形成疏水表面，但是让 RNA 分子精确区分化学结构非常相似的氨基酸几乎是不可能的，例如，Gly 和 Ala 及 Val 和 Ile 之间，其差别仅仅在于一个甲基（CH_3）的存在与否。所以 Crick 提出，在整合成蛋白质之前，氨基酸先结合在某种特异的适配子（adaptor）分子上，而适配子具有独特的表面，可以与 RNA 模板上的碱基特异结合。

转运 RNA 的发现

有发现表明，蛋白质的合成依赖于无细胞体系的建立。加入 RNA 分子，该体系能够以氨基酸前体进行基本的蛋白质合成。1953 年，Paul C. Zamecnik 及其同事最早有效地建立了这种体系。他们的成功主要得益于最新的放射性标记氨基酸的技术，这种技术可以用来跟踪新合成的微量的蛋白质，另外还有高质量的、易于使用的用来分离各种细胞组分的制备超离心技术。在早期，人们就认为蛋白质在细胞内的合成场所是核糖体，核糖体是所有细胞的细胞质中的含 RNA 的小颗粒，它与蛋白质的合成有密切的关系（图 2-13）。

几年后，Zamecnik 与 Mahlon B. Hoagland 合作研究获得了重要的发现：在被整合到蛋白质之前，氨基酸首先结合到我们现在所称的转运 RNA（transfer RNA，tRNA）分子上。转运 RNA 大约占细胞总 RNA 的 10%（图 2-14）。

除 Crick 之外，这一实验结果出乎所有人的预料。Crick 已经预测到所谓的适配子（adaptor）可能是 RNA 短链，因为它们的碱基应该能够与蛋白质合成的模板 RNA 上的相应碱基配对并“阅读”这些碱基。以后我们会讨论更多的细节（第 14 章），实际上 Zamecnik 和 Hoagland 所提出的转运 RNA 分子就是 Crick 提出的适配子分子。每一个 tRNA 分子都有一段连续的碱基序列（反密码子），在蛋白质合成过程中与 RNA 模板上的一组连续碱基序列（密码子）特异结合。

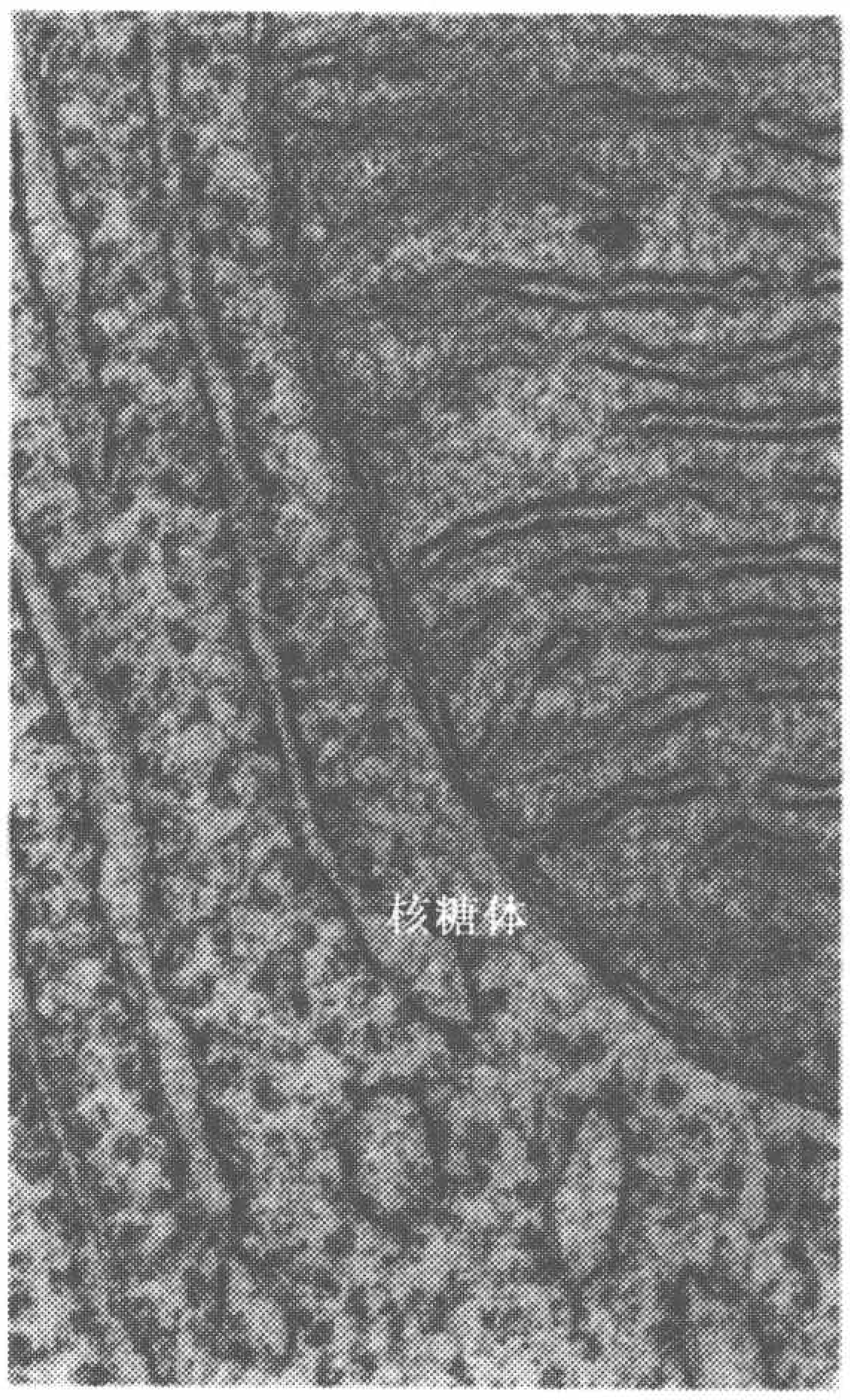


图 2-13　附着在内质网上的核糖体的电子显微照片。 胰腺细胞的电子显微照片（×105 000）。图的右上部分为线粒体，左下部分为附着在内质网上的大量的核糖体，有的核糖体游离存在于细胞质中，有的附着在内质网膜上（K. R. Porter 惠赠）。

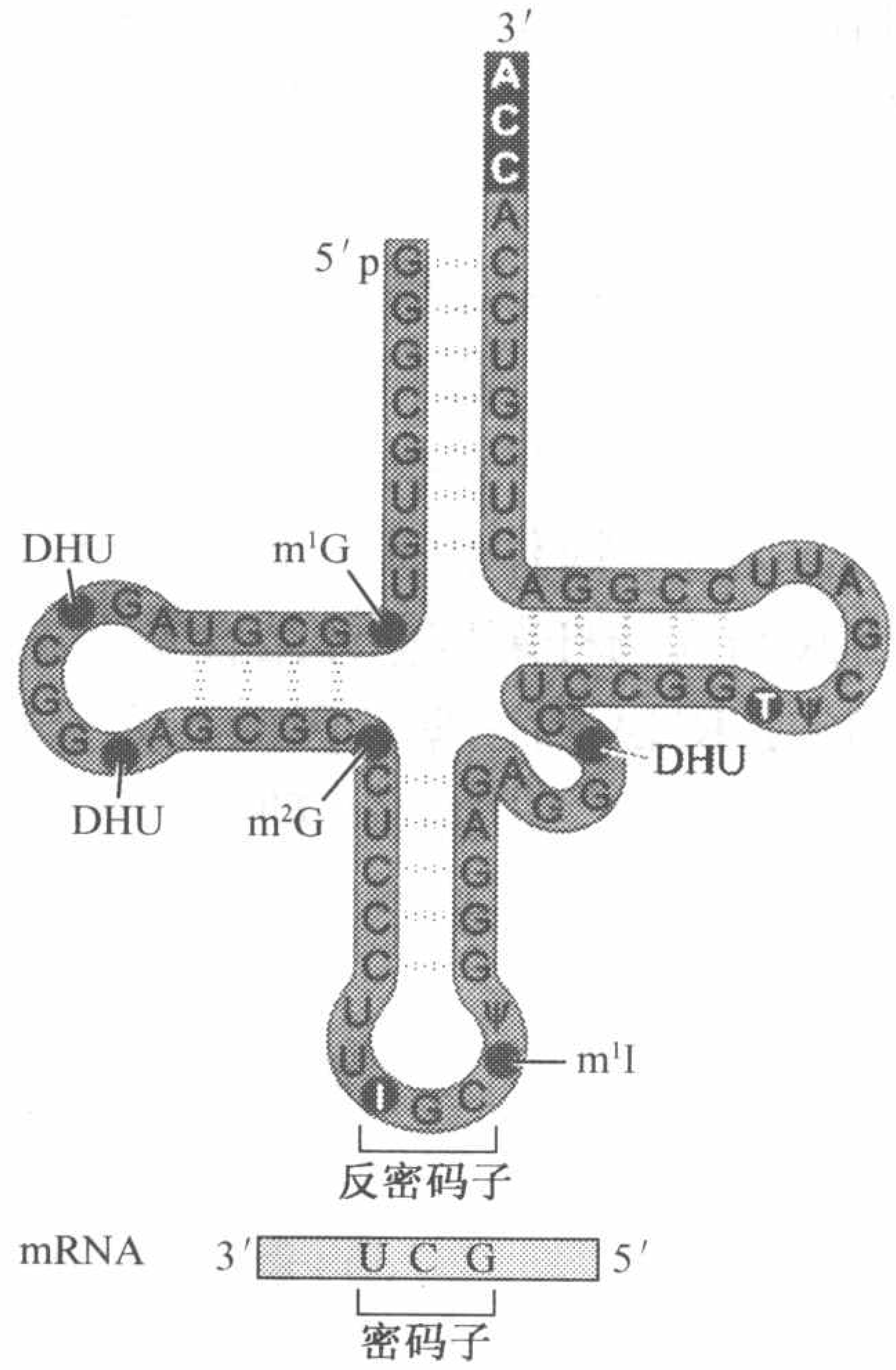


图 2-14　由 Robert W. Holley 及其助手鉴定的酵母丙氨酸 tRNA 的结构。 这个 tRNA 的反密码子识别 mRNA 上丙氨酸。在 tRNA 结构中存在着修饰核苷：ψ＝假尿嘧啶核苷；T＝胸腺嘧啶核糖核苷；DHU＝5，6-二氢尿嘧啶核苷；I＝次黄嘌呤核苷；m¹G＝*N*¹-甲基鸟嘧啶核苷；m¹＝1-甲基次黄嘌呤核苷；m²G＝*N*²-甲基鸟嘧啶核苷。

核糖体的矛盾问题：缺乏特异性

细胞总 RNA 的 85%存在于核糖体中，因为在蛋白质合成旺盛的细胞中（如胰脏、肝脏细胞和快速增殖的细菌），核糖体的绝对数量急剧增加，所以最初人们认为核糖体 RNA（ribosome RNA，rRNA）是决定氨基酸顺序的模板。但当详细研究了大肠杆菌的核糖体之后，发现若干难以解释的现象。首先，所有大肠杆菌核糖体甚至其他所有生物的核糖体，都由两个大小不等的亚基组成，每个亚基都包含 RNA，在周围离子浓度影响下，两个亚基可以可逆地聚合或分开。第二，小亚基中所有 rRNA 的长度大致相同（大肠杆菌约 1500 碱基），大亚基中的 rRNA 长度也基本相同（约 3000 碱基）。第三，在所有已知的细菌、植物和动物中，大小亚基的 rRNA 碱基组成基本相同（有较高的 G 和 C），虽然它们各自 DNA 的 AT/GC 比值有很大变化。如果说 rRNA 是由大量不同基因组成的、不同 RNA 模板的集合体，上述现象都不应该出现，所以大小亚基都不像是模板 RNA。

信使 RNA（mRNA）的发现

T4 噬菌体侵染的细胞提供了一个寻找真正模板的理想体系。受 T4 侵染后，大肠

杆菌细胞停止合成自己的 RNA，只能合成 T4 DNA 转录的 RNA。令人吃惊的是，T4 RNA 不仅与 T4 DNA 的碱基组成很相似，而且它不能与核糖体蛋白结合，正常情况下，核糖体蛋白与 rRNA 结合形成核糖体。相反，T4 RNA 首先与已存在的核糖体结合，然后移动穿过核糖体表面，使其碱基进入适当位点，在这里与相应的 tRNA-氨基酸前体结合进行蛋白质合成（图 2-15）。这种情况下，T4 RNA 就决定了氨基酸的顺序，它就是长期寻找的蛋白质合成的 RNA 模板。由于它将 DNA 信息带到了进行蛋白质合成的核糖体位点，故称为信使 RNA（message RNA，mRNA）。首次观察到 T4 RNA 与大肠杆菌核糖体结合是在 1960 年春，很快在未侵染的大肠杆菌的细胞中获得了其他信使 RNA 的证据，因此，完全排除了 rRNA 作为模板的可能。实际上，核糖体中的 rRNA 组分，与大约 50 种与之结合的核糖体蛋白一起，是蛋白质合成的工厂将 tRNA-氨基酸前体引导到某个位置，在那里它们读取 mRNA 模板提供的信息（详见第 14 章）。

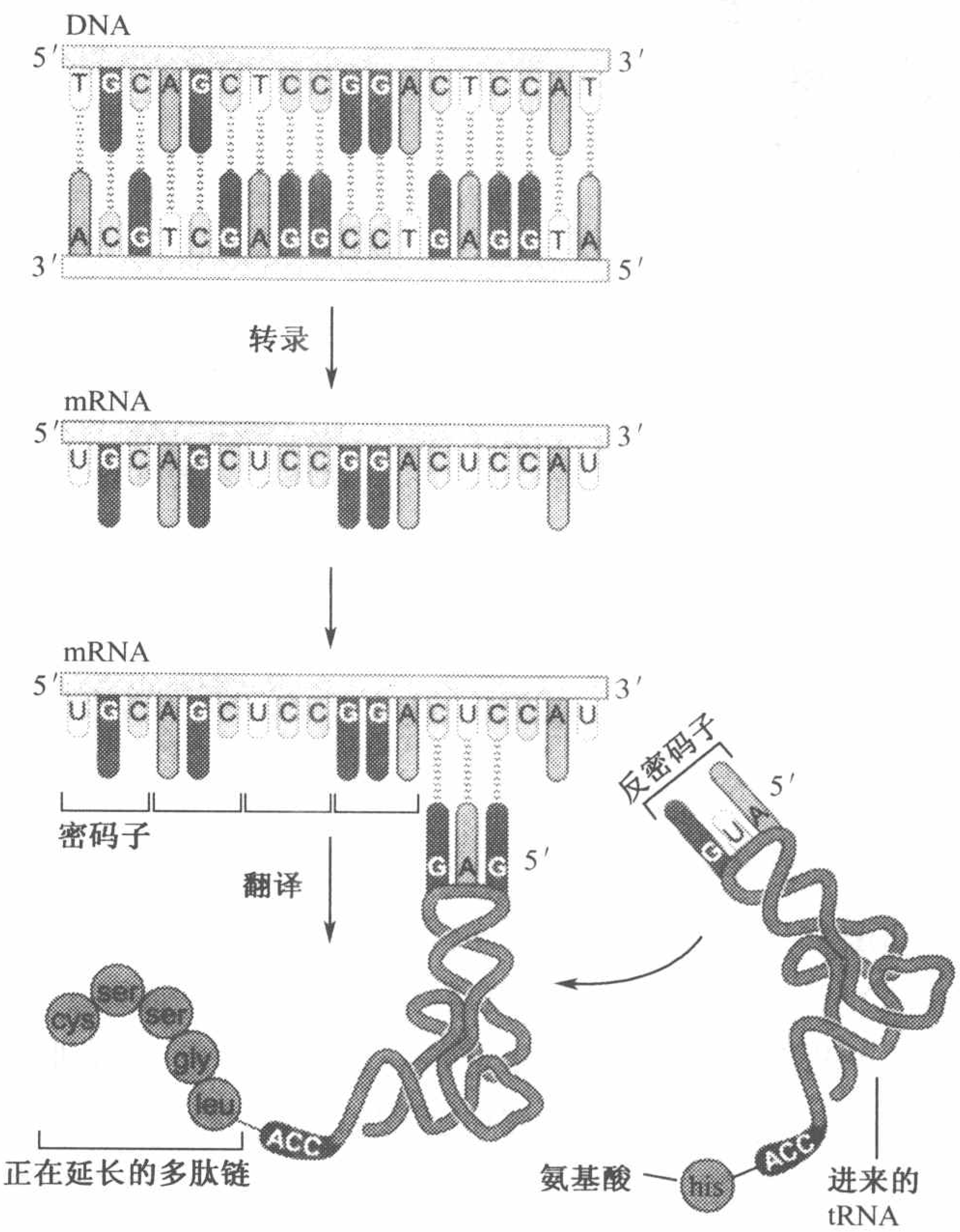


图 2-15 转录和翻译。以 DNA 的一条链为模板，按照碱基互补原则合成 mRNA。密码子每 3 个一组，与 tRNA 分子上反密码子区域的 3 个核苷互补。碱基配对形成后，tRNA 分子另一端携带的氨基酸添加到正在生长的蛋白质链上。

细胞中，mRNA 仅占总 RNA 的百分之几。为编码细胞里的不同的蛋白质，mRNA 的长度和核苷酸组成变化很大，所以，不难理解为什么 mRNA 在最初未被研究者们重视。在给定的时间内，只有很小的一段 mRNA 结合在核糖体上，所以一个 mRNA 分子可以被

多个核糖体同时读取。绝大多数核糖体都是多聚核糖体（polyribosome）（翻译相同 mRNA 的一组核糖体）的一部分，多聚核糖体可能会包含 50 多个成员（图 2-16）。

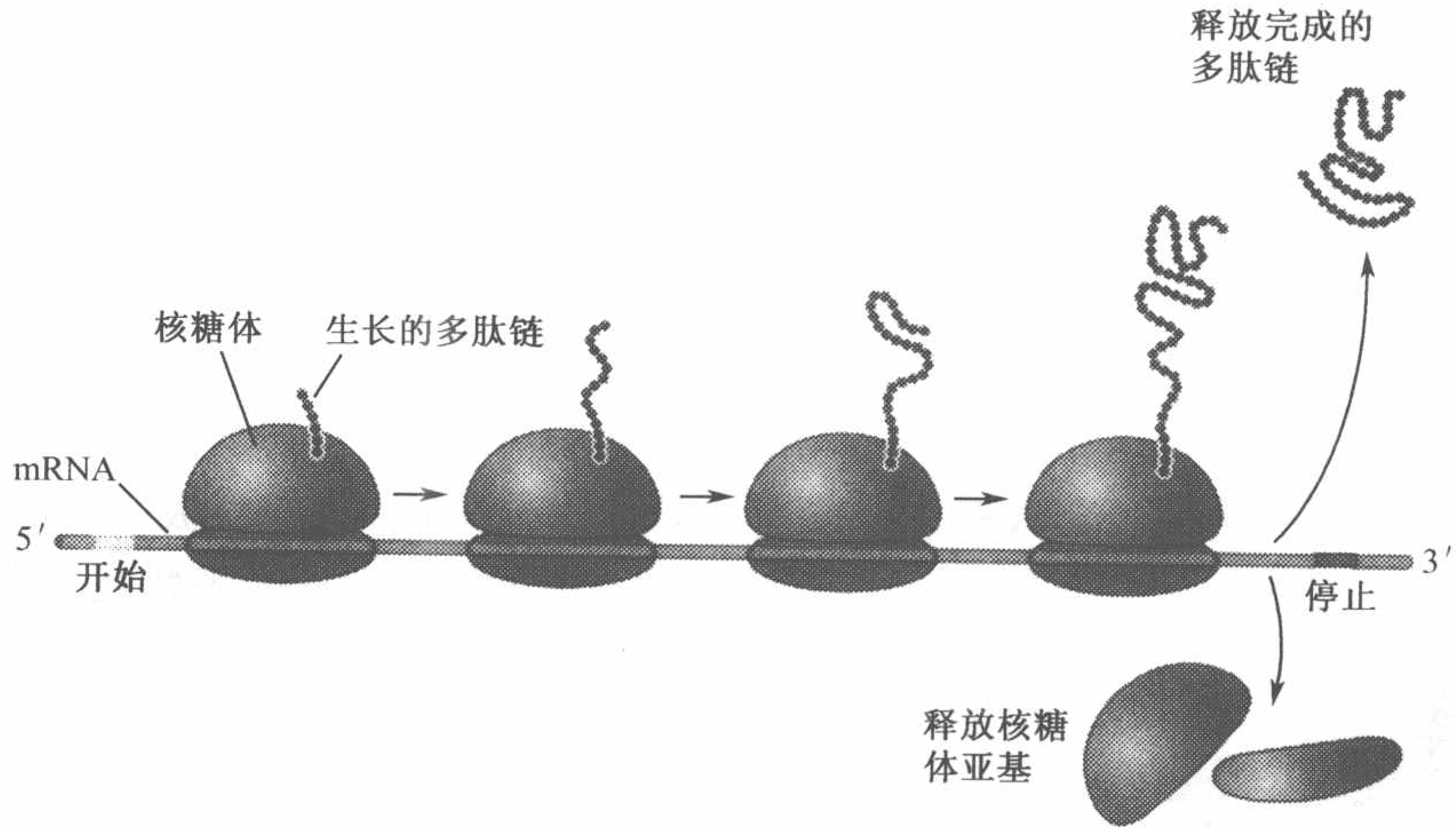


图 2-16　多聚核糖体示意图。每个核糖体附着在 mRNA 5′端的信号起始处，并沿着 mRNA 分子合成多肽。一条 mRNA 链上可能同时结合数个核糖体，整个装置称为多聚核糖体。

以 DNA 为模板酶促 RNA 的合成

随着信使 RNA 的发现，第一个能以 DNA 为模板转录 RNA 的酶很快就分别在生化学家 Jerard Hurwitz 和 Sam B. Weiss 的实验室分离出来了。这种称为 RNA 聚合酶（RNA polymerase）的酶必须在有 DNA 的情况下工作，并以其为模板，利用 ATP、GTP、CTP 和 UTP 核苷酸为前体，合成单链 RNA（图 2-17）。在细菌中，同一种酶分别以相应的染色体 DNA 为模板合成 RNA。生成的 RNA 碱基组成随加入不同 AT/GC 比值的 DNA 模板而变化，这就是 DNA 负责排列核糖核苷酸前体的直接证据。在所有的酶促合成中，RNA 的 AU/GC 比值与 DNA 的 AT/GC 比值大体相同（表 2-2）。

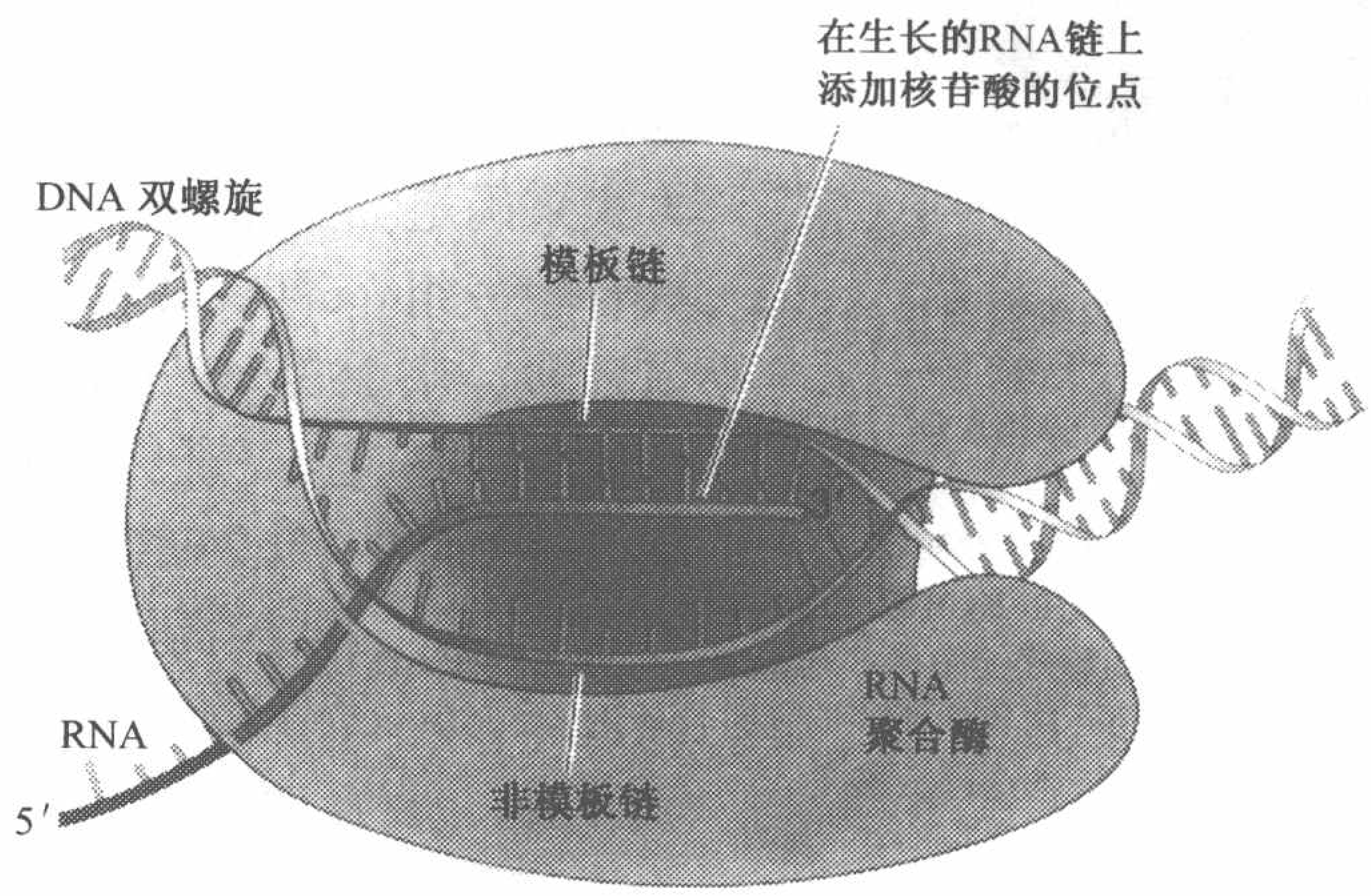


图 2-17　RNA 聚合酶催化的以 DNA 为模板的 RNA 的合成。

表 2-2 酶促合成 RNA 与其双螺旋 DNA 模板碱基组成的比较

DNA 模板来源	RNA 碱基组成				A+U/G+C	A+T/G+C
	腺嘌呤	尿嘧啶	鸟嘌呤	胞嘧啶	观察	DNA 中
T2	0.31	0.34	0.18	0.17	1.86	1.84
小牛胸腺	0.31	0.29	0.19	0.21	1.50	1.35
大肠杆菌	0.24	0.24	0.26	0.26	0.92	0.97
Micrococcus lysodeikticus（一种细菌）	0.17	0.16	0.33	0.34	0.49	0.39

在转录过程中，两条 DNA 链中只能有一条作为 RNA 合成的模板。这是有道理的，因为两条 DNA 链虽然序列互补但携带的遗传信息并不相同，会编码完全不同的多肽。RNA 的合成总是沿固定方向进行，由 5′端的核苷酸开始，到 3′端的核苷酸结束（图 2-17）。

到此为止，已经有充分的证据证明了所推测的 RNA 确实是从含 DNA 的核中移动到含核糖体的细胞质中。将细胞短暂地暴露在放射性标记的前体中，然后加入过量的未标记氨基酸（“脉冲标记”实验），在一段很短的时间内合成的 mRNA 就会被标记。这项研究表明 mRNA 是在细胞核中合成的。一个小时内，就可以观察到大部分 RNA 离开了细胞核进入细胞质中（图 2-18）。

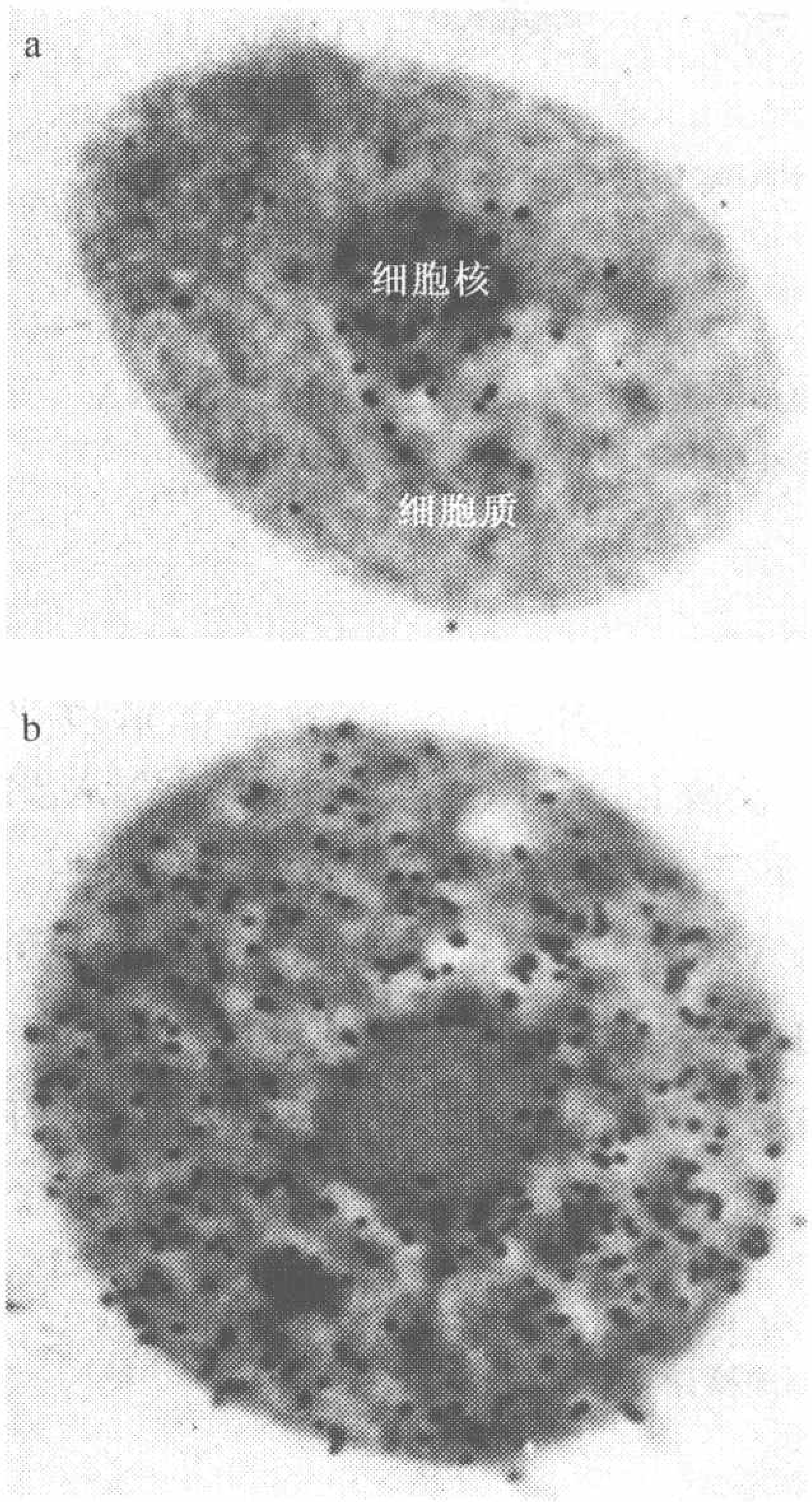


图 2-18 RNA 在核中合成后转移到细胞质中的证明。 a. 在放射性标记的胞嘧啶培养基中培养 15min 后细胞（四膜虫）的放射自显影照片。细胞薄切片照片中深度曝光的是银乳胶颗粒，每个黑点代表掺入到 RNA 中的 ^{3}H发射电子的轨迹，几乎所有新形成的 RNA 都在细胞核中。b. 相同的细胞在放射性标记的胞嘧啶中培养 12min，然后转移到非放射性标记的胞嘧啶中培养 88min 后的放射自显影照片。发现在 12min 内掺入 RNA 中的所有的标记，都离开细胞核进入到细胞质中。（D. M. Prescott，University of Colorado Medical School；reproduced from Prescott D. M. 1964. Progr. Nucleic Acid Res Mol Biol. 3：35. © Elsevier。经许可）

遗传密码的破译

已知存在 20 种氨基酸，但仅有 4 种碱基，所以肯定是由几个核苷酸的组合以某种方式来确定一个特定的氨基酸。如果是两个核苷酸一组，最多只能定义 16 个（4×4）氨基酸。从 1954 年开始，人们就在认真考虑遗传密码应该是什么样的问题，尤其是三联体（3 个核苷酸一组）会如何行使功能，即使三联体显然能提供多于需要的排列组合（4×4×4）（如果每个氨基酸只能由一个三联体编码）。此时提出了共线性假设，这是非常重要的。共线性的意思是沿 DNA 链排列的核苷酸组合编码沿多肽链上排列的氨基

酸。事实上，Charles Yanofsky 和 Sydney Brenner 在 20 世纪 60 年代初期就已经通过对细菌蛋白质的突变分析观察到这种共线性的存在。另外，还有 Brenner 和 Crick 在 1961 年的遗传分析也同样重要，他们首次证明了 3 个核苷酸一组确定一个氨基酸。

但是哪 3 个碱基的特定组合（密码子）确定哪一种氨基酸只有通过生化分析才能知道。突破性的进展来自于 Marshall Nirenberg 和 Heinrich Mattaei 的工作。1961 年他们在一起工作时观察到：将体外合成的多聚核苷酸 poly U（UUUUU…）加入能合成蛋白质的无细胞体系中，产生的是只含苯丙氨酸的多肽。所以，UUU 这一组核苷酸肯定编码的是苯丙氨酸。利用逐渐复杂的、确定的多聚核苷酸作为信使 RNA，很快就鉴定出了越来越多的密码子。在破译密码子的过程中，尤为重要的是利用了像 AGUAGU 这样的多聚核苷酸，这是由有机化学家 Har Gobind Khorana 合成的。这些特定的多聚核苷酸在破译更复杂的密码子中是至关重要的。1966 年，密码子全部破译，可能的 64 个密码子中有 61 个编码相应的氨基酸，大多数氨基酸都是由一个以上的核苷酸三联体编码的（表 2-3）。

表 2-3 遗传密码

第1个位置	第2个位置 U	第2个位置 C	第2个位置 A	第2个位置 G	第3个位置
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
U	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
U	UUA Leu	UCA Ser	UAA 终止子	UGA 终止子	A
U	UUG Leu	UCG Ser	UAG 终止子	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
C	CUC Leu	CCC Pro	CAC His	CGC Arg	C
C	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
C	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
A	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
A	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A
A	AUG Met	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
G	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
G	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
G	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

确定蛋白质合成的方向

确定了遗传密码的性质之后，下一个的问题就是多聚核苷酸链如何指导多肽的合成。正如我们在本章已经看到，在第 6 章还要进行详细讨论的，多聚核苷酸链（DNA 和 RNA）由 5′→3′方向合成。但是多肽链的延伸是如何进行的呢？是从氨基端到羧基

端，还是相反呢？

利用与确定氨基酸密码相似的无细胞体系蛋白质合成系统回答了上述问题。研究者加入 β 珠蛋白 mRNA，而不是合成的 mRNA，来指导蛋白质的合成。在蛋白质合成开始几分钟后，利用放射性标记的氨基酸处理无细胞体系几秒钟（低于合成珠蛋白链需要的时间），然后立即终止蛋白质合成。这种快速的放射性标记方式称之为脉冲（pulse）或脉冲标记（pulse-labeling）。利用凝胶电泳技术将在标记期间内合成的完整的珠蛋白与未完全合成的珠蛋白分离（第 20 章），分析所有的在完成合成时都存在放射性标记的蛋白质，用胰蛋白酶处理全长的多肽，使蛋白质在特异位点断裂，产生一系列长度不同的多肽片段，最后，测量掺入每个片段中的放射性强度（图 2-19）。

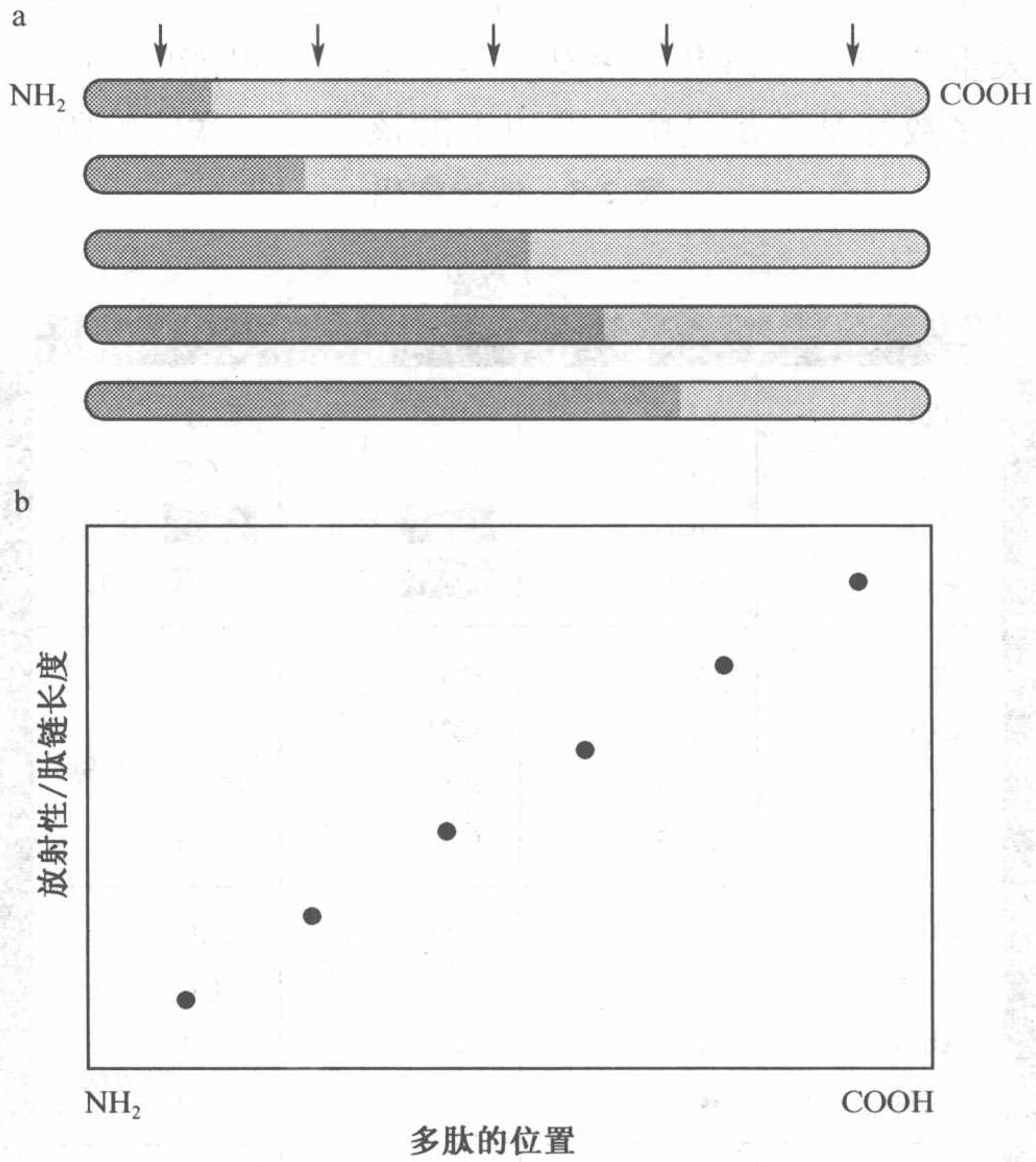


图 2-19 掺入到延伸的多肽链的放射性标记。 a. 脉冲标记后合成多肽链中的放射性分布（示蓝色。β 珠蛋白的胰蛋白酶切割位置以红色箭头表示。）b. 放射性掺入对多肽的位置作图。

注意珠蛋白是在脉冲标记的不同阶段合成的最后合成的肽链含有最多的放射性标记的前体。（图 2-19a）。相反，含放射性标记最弱的肽链来自最早合成的 β 珠蛋白。研究者观察到来自珠蛋白的氨基端区域的肽链的放射性最低，而羧基端区域的标记最强（图 2-19b）。

所以我们得出如下结论：蛋白质的合成方向是从氨基端到羧基端，换句话说，在蛋白质合成过程中，新的氨基酸是加到新生多肽链的羧基末端上的。

起始和终止信号也由DNA编码

开始人们推测mRNA的翻译是从一端开始，读完整个mRNA信息之后结束。但实际上翻译的起始和终止都在mRNA的内部位置上，所以翻译的起始和终止信号肯定存在于DNA（及其mRNA产物）上。首先被阐明的是终止信号，3个密码子（UAA、UAG和UGA），即已知的无义密码子（nonsense codon），不编码任何一个氨基酸，而是作为翻译的终止信号（有时称为终止密码子）。翻译的起始信号的编码方式较复杂，所有多肽链均以甲硫氨酸开始，但是三联体密码子AUG既编码起始的甲硫氨酸，也编码内部的甲硫氨酸。原核生物中，在编码多肽起始的AUG密码子前面有富含嘌呤核苷酸的区域，其作用是使mRNA结合到核糖体上（第14章）。在真核生物中，AUG与mRNA起始点的相对位置起关键作用，而第一个AUG总是被选为翻译的起初位置。

基因组学时代

随着中心法则的阐明，到20世纪60年代中期，对于蕴含在核苷酸序列内的遗传蓝图决定表型这一问题有了清晰的认识，这意味着从DNA序列中可能揭示出对生物及其进化本质的深刻认识。近年来由于快速、自动化DNA测序方法的出现，几百种有机体的全基因组序列测定工作已经完成，甚至单个拷贝就达30亿碱基对的人类基因组序列，及其包含的不少于20 000个的基因也已经阐明。

多种生物的基因组测序使基因组序列的比较分析的用途越来越重要。通过比较不同生物相似基因编码的氨基酸预测序列，人们可以鉴别出重要的蛋白质区段。例如，在许多不同的生物中，负责核苷酸结合或直接催化核苷酸添加的DNA聚合酶的氨基酸序列是非常保守的。所以，鉴别在细菌中保守的、对DNA聚合酶功能至关重要的氨基酸序列只能通过对细菌的基因组序列的比较获得，而在真核细胞中不一定保守。

比较不同的基因组也能了解非编码蛋白质的DNA序列。对控制基因表达、DNA复制、染色体分离和重组的序列的鉴别来说，进行基因组序列的比较都会有很大的帮助。因为这些调控序列更倾向于以较快的速度发生变化，这样的比较往往是在亲缘关系较近的物种中（如不同的细菌之间或人与其他灵长类之间）

因为对亲缘关系较近的物种进行比较具有重要的价值，所以促进了与研究得较多的模式生物（如果蝇、酵母和多种灵长类动物）。亲缘关系较近的生物的基因组测序工作的开展，对同一生物不同个体的基因组进行比较，对鉴别疾病相关的突变具有潜在的价值。如，为了鉴别疾病相关基因，最近建立了可以快速地对许多不同个体基因组片段进行比较的方法。最后，我们可以预想有一天，通过比较基因组序列分析揭示人类复杂行为的起源，如语言的获得，还有动物身躯进化的多样性的机制等。

以后的章节将为理解DNA如何作为模板、决定生物学复杂性打下坚实的基础。第1篇的剩余章节用来回顾与本书主题相关的基本的化学和生物学内容；第2篇基

因组的维持，叙述遗传物质的结构及其忠实的复制；第 3 篇基因组的表达，描述 DNA 中的遗传指令是如何转变为蛋白质的；第 4 篇调控，讲述产生机体复杂性（如胚胎发生学）和物种多样性（如进化）的不同基因表达策略；最后，第 5 篇方法学，介绍各种实验室技术，生物信息学方法和普遍用于生物学问题研究的模式系统。

小结

DNA 是遗传物质的发现可以追溯到 Griffith 的实验，他发现细菌的无毒菌株能被一种加热致死的有毒菌株中的物质遗传转化。随后 Avery、McCarty 和 MacLeod 证实此转化物质是 DNA。Hershey 和 Chase 用放射物质标记的噬菌体进行实验，进一步证实 DNA 是遗传物质。在 Chargaff 定律及 Franklin 和 Wilkins 的 X 射线衍射研究的基础上，Watson 和 Crick 提出了 DNA 的双螺旋结构。在这个模型中，两条多肽链相互缠绕形成规则的双螺旋，双螺旋的两条链通过碱基对之间的氢键而联系。腺嘌呤总是与胸腺嘧啶配对，鸟嘌呤与胞嘧啶配对。碱基对的存在说明两条链的序列不同但互补。这种互补的双链结构意味着一种复制机制，一条链可以作为模板合成其互补链。支持这一假说的证据，一是 Meselson 和 Stahl 观察到的每一次 DNA 复制过程中双螺旋的两条链的相互分离，二是 Kornberg 发现了使用单链 DNA 作为模板合成其互补链的酶。

就像我们看到的，按照中心法则，遗传信息从 DNA 流到 RNA 再到蛋白质，这一传递过程分两步完成。首先 DNA 转录为 RNA 中间物（mRNA），然后再翻译成蛋白质，mRNA 的翻译需要有叫做 tRNA 的 RNA 适配子。遗传密码的关键特征是每一个三联体密码被 tRNA 识别，tRNA 可以携带相应的氨基酸。在 64 个（4×4×4）可能的密码子中，61 个用于编码构成各种蛋白质的 20 种氨基酸，另外 3 个用于提供终止信号。对遗传密码的认识使我们从 DNA 序列能预测蛋白质序列。快速 DNA 测序方法的出现使我们进入了新的基因组时代，很多生物有机体包括人类的全基因组序列正在完成基因组序列的比较为鉴别基因组中的重要区段提供了一个有效的方法，不仅是编码重要的蛋白质元件，还有那些控制基因表达和基因组复制的调控区域。测序正在进行。

（刘国振 译 吴东颖 刘国振 刘韧 杨焕明 校）

参考文献

Brenner S., Jacob F., and Meselson M. 1961. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* **190:** 576–581.

Brenner S., Stretton A.O.W., and Kaplan S. 1965. Genetic code: The nonsense triplets for chain termination and their suppression. *Nature* **206:** 994–998.

Cairns J., Stent G.S., and Watson J.D., eds. 1966. *Phage and the origins of molecular biology.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Chargaff E. 1951. Structure and function of nucleic acids as cell constituents. *Fed. Proc.* **10:** 654–659.

Cold Spring Harbor Symposia on Quantitative Biology. 1966. Vol. 31: *The genetic code.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Crick F.H.C. 1955. On degenerate template and the adaptor hypothesis. A note for the RNA Tie Club, unpublished. Mentioned in Crick's 1957 discussion, pp. 25–26, in The structure of nucleic acids and their role in protein synthesis. *Biochem. Soc. Symp.* no. 14, Cambridge University Press, Cambridge, England.

———. 1958. On protein synthesis. *Symp. Soc. Exp. Biol.* **12:** 548–555.

———. 1963. The recent excitement in the coding problem. *Prog. Nucleic Acid Res.* **1:** 164–217.

———. 1988. *What mad pursuit: A personal view of scientific discovery.* Basic Books, New York.

Crick F.H.C. and Watson J.D. 1954. The complementary structure of deoxyribonucleic acid. *Proc. Roy. Soc. A* **223:** 80–96.

Echols H. and Gross C.A., eds. 2001. *Operators and promoters: The story of molecular biology and its creators.* University of California Press, Berkeley, California.

Franklin R.E. and Gosling R.G. 1953. Molecular configuration in sodium thymonuclease. *Nature* **171:** 740–741.

Hershey A.D. and Chase M. 1952. Independent function of viral protein and nucleic acid on growth of bacteriophage. *J. Gen. Physiol.* **36:** 39–56.

Hoagland M.B., Stephenson M.L., Scott J.F., Hecht L.I., and Zamecnik P.C. 1958. A soluble ribonucleic acid intermediate in protein synthesis. *J. Biol. Chem.* **231:** 241–257.

Holley R.W., Apgar J., Everett G.A., Madison J.T., Marquisse M., Merrill S.H., Penswick J.R., and Zamir A. 1965. Structure of a ribonucleic acid. *Science* **147:** 1462–1465.

Ingram V.M. 1957. Gene mutations in human hemoglobin: The chemical difference between normal and sickle cell hemoglobin. *Nature* **180:** 326–328.

Jacob F. and Monod J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3:** 318–356.

Judson H.F. 1996. *The eighth day of creation*, expanded edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Kornberg A. 1960. Biological synthesis of deoxyribonucleic acid. *Science* **131:** 1503–1508.

Kornberg A. and Baker T.A. 1992. *DNA replication.* W.H. Freeman, New York.

McCarty M. 1985. *The transforming principle: Discovering that genes are made of DNA.* Norton, New York.

Meselson M. and Stahl F.W. 1958. The replication of DNA in *Escherichia coli. Proc. Natl. Acad. Sci.* **44:** 671–682.

Nirenberg M.W. and Matthaei J.H. 1961. The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci.* **47:** 1588–1602.

Olby R. 1975. *The path to the double helix.* University of Washington Press, Seattle.

Portugal F.H. and Cohen J.S. 1980. *A century of DNA: A history of the discovery of the structure and function of the genetic substance.* MIT Press, Cambridge, Massachusetts.

Sarabhai A.S., Stretton A.O.W., Brenner S., and Bolte A. 1964. Co-linearity of the gene with the polypeptide chain. *Nature* **201:** 13–17.

Stent G.S. and Calendar R. 1978. *Molecular genetics: An introductory narrative*, 2nd ed. Freeman, San Francisco.

Volkin E. and Astrachan L. 1956. Phosphorus incorporation in *E. coli* ribonucleic acid after infection with bacteriophage T2. *Virology* **2:** 146–161.

Watson J.D. 1963. Involvement of RNA in synthesis of proteins. *Science* **140:** 17–26.

———. 1968. *The double helix.* Atheneum, New York.

———. 1980. *The double helix: A Norton critical edition* (ed. G.S. Stent). Norton, New York.

———. 2000. *A passion for DNA: Genes, genomes and society.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

———. 2002. *Genes, girls, and Gamow: After the double helix.* Knopf, New York.

Watson J.D. and Crick F.H.C. 1953a. Genetical implications of the structure of deoxyribonucleic acid. *Nature* **171:** 964–967.

——— 1953b. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171:** 737–738.

Wilkins M.H.F., Stokes A.R., and Wilson H.R. 1953. Molecular structure of deoxypentose nucleic acid. *Nature* **171:** 738–740.

Yanofsky C., Carlton B.C., Guest J.R., Helinski D.R., and Henning U. 1964. On the colinearity of gene structure and protein structure. *Proc. Natl. Acad. Sci.* **51:** 266–272.

第3章 弱化学作用的重要性

本章概要

•

化学键的特征

•

自由能的概念

•

生物体系中的弱化学键

本书重点描述的，也是大多数分子生物学家所关注的，是蛋白质和核酸。它们分别由氨基酸和核苷酸构成，并且两者都是由于共价键的参与而组成了多肽（蛋白质）和多聚核苷酸（核糖核酸）。共价键是一种强的、稳定的化学键，在生物体系中一般不会自发断裂。但是，弱一些的化学键也同样存在，实际上，它们对生物体是至关重要的，这部分是由于它们能够在细胞内生理环境下形成或断裂。弱化学键介导了酶与其底物的相互作用，更重要的一点是，弱化学键介导了大分子之间的相互作用（将在后面几章中讨论）。例如，蛋白质与DNA或RNA，一种蛋白质与另一种蛋白质之间的相互作用。然而，同等重要的是，弱化学键同样介导了同一大分子不同部位之间的相互作用，并由此决定了这些分子的形状及其生物学功能。因此，尽管蛋白质是由许多共价连接的氨基酸组成的线性单链，但是它的形状和功能是由它形成的稳定的三维结构所决定的。即使是在蛋白质一级顺序中不相邻的两个氨基酸之间也存在弱相互作用，这些各自独立的弱相互作用的总和最终决定了蛋白质的形状。同样地，正是这些弱的非共价键使DNA的两条链聚合在一起形成双螺旋结构。

本章将要讨论化学键的本质，重点是讨论对所有生物大分子的正确功能起着至关重要作用的弱化学键的本质。我们将着重介绍弱化学键何以拥有其特性。这一类化学键包括范德华力、疏水键、氢键以及离子键。

化学键的特征

化学键（chemical bond）是一种将原子联系在一起的吸引力。有限数量原子的集合体称为分子。起初，人们认为在分子中只有共价键将原子联系在一起；现在，我们已经知道弱吸引力在许多大分子的联系中起到重要作用。例如，一些弱化学键的联合作用将血红素的4条多肽链联系起来。这就是为什么习惯上弱相互作用也被称为化学键，即使弱相互作用相当弱，以至于当其单独存在时不能有效连接两个原子。化学键有几个共同特征，一个显而易见的特征就是强度。在生理温度下，强化学键几乎从不断裂，这就

是为什么共价键连接的原子总是属于同一个分子；弱化学键则比较容易断裂，单个只能短暂存在。只有当众多弱化学键组成有序的集合时，它们才能存在较长时间。化学键的强度与它的键长有关，所以由强化学键连接的两个原子之间的距离通常比由弱化学键连接的两个原子之间的距离短。例如，构成氢分子（H：H）的两个共价连接氢原子相距0.74Å，而由范德华力连接的两个同样的氢原子则相距1.2Å。

另一个重要的特征是原子能形成的最大键数。一个原子能够形成的共价键的数目称为原子价（valence)。例如，氧的原子价为2，即它不能形成两个以上的共价键。而对于范德华力，由于其限制因素纯粹是空间效应，因此有更大可变性，可能成键的数目仅由同时能够接触到的相邻原子的数目决定。氢键的形成有更严格的限制，一个共价连接的氢原子通常只能参与一个氢键的形成，而氧原子在少数情况下可以形成两个以上的氢键。

由同一个原子起始的两个化学键之间的角度称为键角（bond angle)。两个特定共价键之间的键角总是一致的。例如，一个碳原子拥有4个共价键，它们构成正四面体（键角＝109°）；相反，弱化学键的键角具有更大的可变性。

化学键的允许旋转自由度（freedom of rotation）也是各不相同。单个共价键允许连接的原子自由旋转（图3-1），而二键和三键就有着严格限制。具有部分双键特性的共价键，如肽键，也有着严格限制。因此，肽键连接起来的羰基（C═O）和亚氨基(N═C)基团必须处在同一平面（图3-2)。另一方面，更弱的离子键对连接的原子的相对方向没有限制。

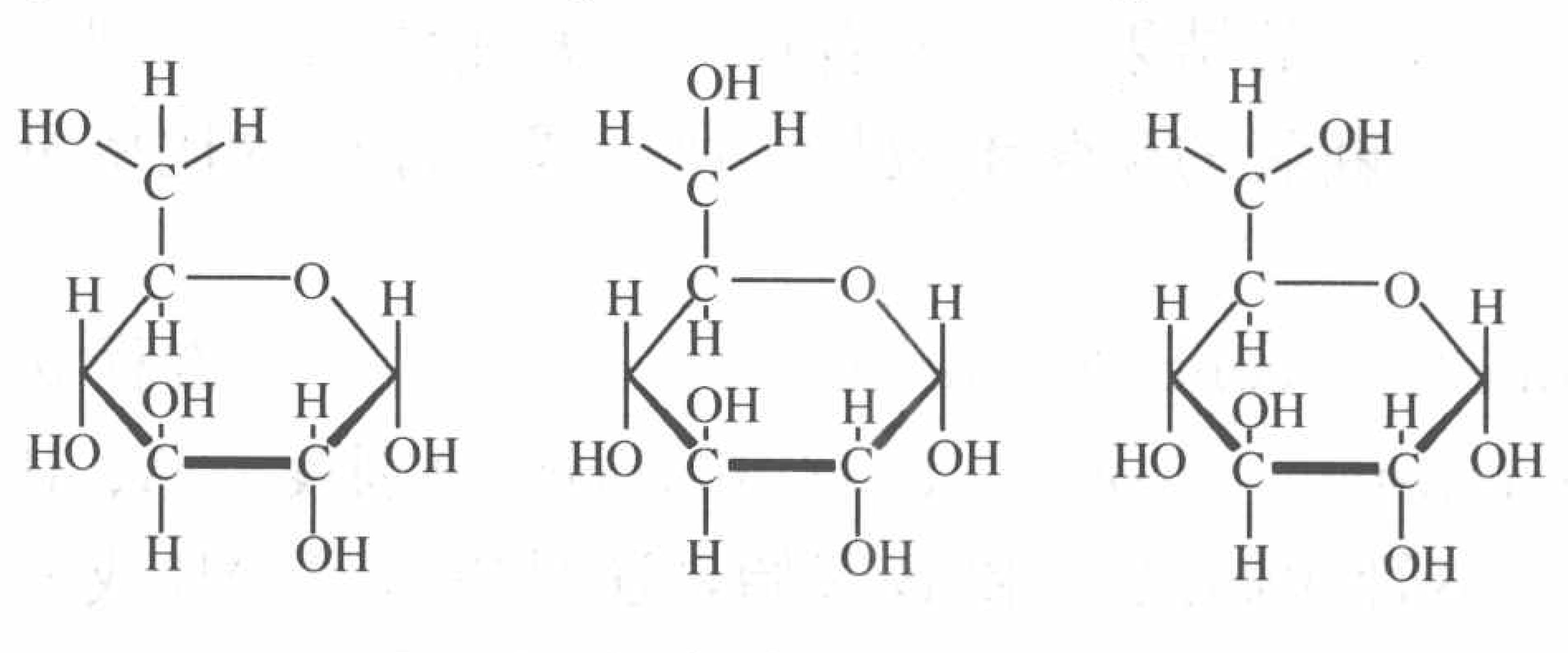


图3-1 葡萄糖中 C_5-C_6 键的旋转。这个碳碳键是一个单键，所以a、b、c所列3种结构都可能存在。

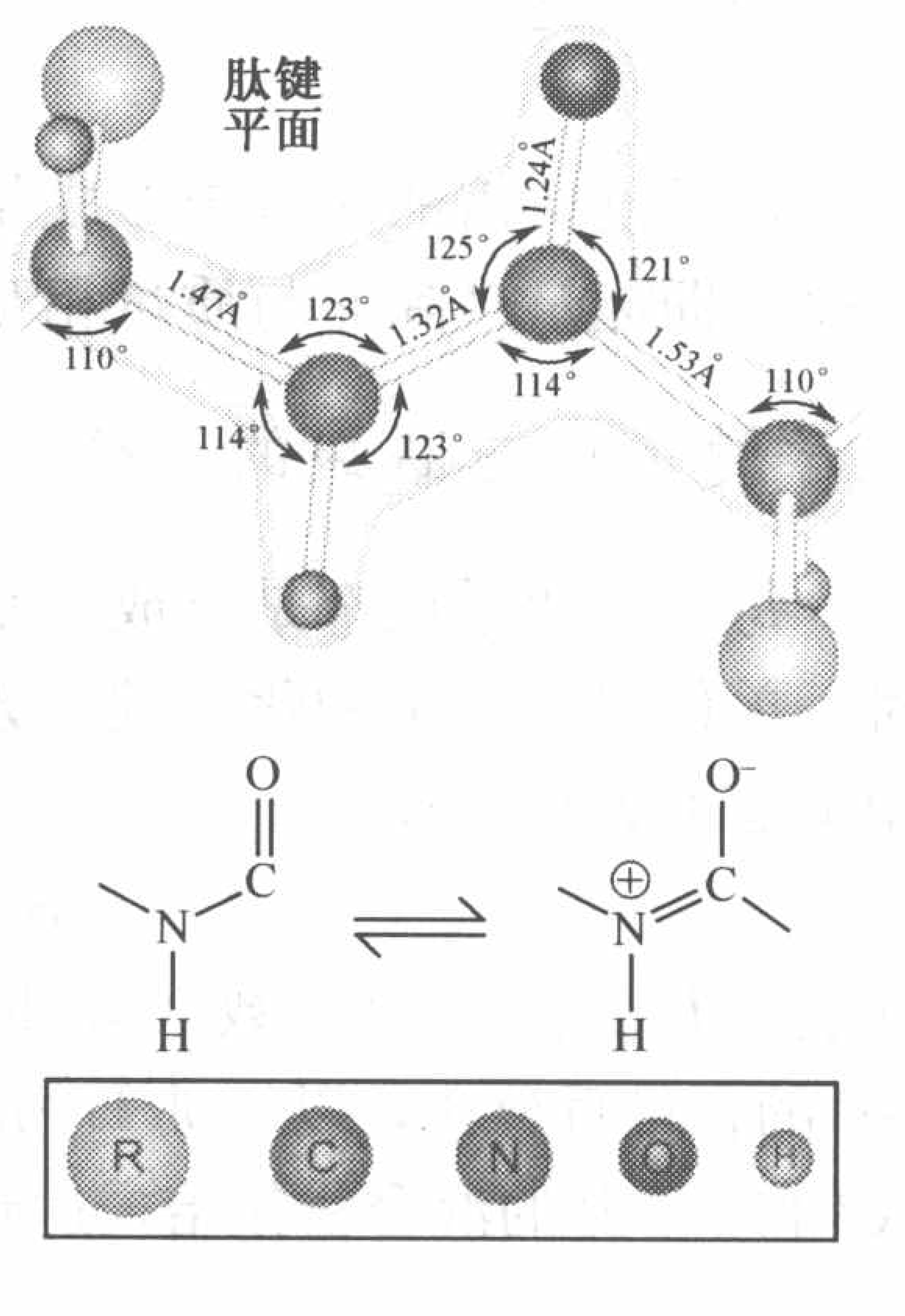


图3-2 肽键的平面结构。上图是某延伸多肽链的片段。由于肽键的双键特性（中间部分)，旋转是不允许的，所有灰色区域的原子必须处于同一平面。然而，围绕其余两个共价键的旋转是可能的，这导致了多肽的不同结构。（来源：Pauling L. 1960. The nature of the chemical bond and the structure of molecules and crystals: An introduction to modern structure chemistry, 3rd edition, p. 495. Copyright© 1960 Cornell University.）

化学键的量子机制解释

构成化学键的力，不论强弱，其本质对化学家来说始终是个谜，直到20世纪20年代，原子的量子理论（量子机制）建立，关于化学键形成的众多的经验主义法则才第一次有了坚实的理论基础。人们认识到所有的化学键，不论强弱，均基于电荷力。量子机制将共价键解释为共用电子，也为弱化学键的形成提供了解释。

化学键的形成涉及能量形式的变化

两个原子之间化学键的自发形成过程总是包括未成键原子部分内能的释放和向另一能态的转变。键越强，成键过程中释放的能量越多。原子 A 和 B 的成键过程可描述为：

$$A+B \rightarrow AB+能量 \quad [方程\ 3\text{-}1]$$

在这里 AB 代表成键集合体，A 与 B 的反应速率直接与两者的碰撞概率成正比。卡(cal) 经常用来作为衡量能量的单位，1cal 是 1g 水的温度从 14.5℃提高到 15.5℃所需的能量。由于 1mol 化学键的打断通常会释放上千卡的能量，因此化学反应中大部分的能量变化用千卡每摩尔表示。

然而，由化学键连接的原子不会永久结合在一起，因为存在着打断化学键的力。到目前为止，其中最重要的一种来自热能。高速运动的分子或原子之间的碰撞能够打断化学键，在碰撞过程中，一个运动分子碰到两个连接着的原子，失去部分动能，同时会将它们分开。分子运动越快（温度越高），在碰撞过程中打断化学键的可能性越大。因此，提高分子所处的温度，其化学键的稳定性就下降。化学键的打断一般表示如下：

$$AB+能量 \rightarrow A+B \quad [方程\ 3\text{-}2]$$

打断化学键所需的能量与形成化学键时释放的能量正好相等。这种等量遵循热力学第一定律，即能量（除非与质量相互转换）既不能创生也不能消灭。

化学键形成与断裂的平衡

每一个化学键都是形成与断裂综合反应的结果。当两者在一个密闭体系中达到平衡时，单位时间产生的化学键的数量将等于断裂的化学键数量。所以成键原子的比率可以用下面的方程描述：

$$K_{eq}=conc^{AB}/(conc^{A}\times conc^{B}) \quad [方程\ 3\text{-}3]$$

在这里 K_{eq}表示平衡常数（equilibrium constant），$conc^{A}$、$conc^{B}$ 和 $conc^{AB}$分别表示 A、B 和 AB 的含量百分比，以 mol/L 为单位。不管我们以 A 和 B 起始，还是以分子 AB 或分子 AB 和 A、B 的混合物起始，达到平衡时，A、B 和 AB 的比率都会达到由 K_{eq}给出的浓度。

自由能的概念

在成键原子的比率趋向平衡浓度的过程中，能量的形式在不断变化。从生物学的角度来说，表示这种能量变化的最有效方法就是通过物理化学家的自由能（free energy）的概念来描述。这一概念用 G 表示，用以纪念 19 世纪伟大的物理学家吉布斯（Josiah Gibbs）。本书不会给出严格的自由能的定义，也不会给出自由能与其他能量形式的区别之处。因此，读者必须阅读有关热力学第二定律的化学类参考书。在此我们只能说，自由能是一种能够做功的能量。

热力学第二定律告诉我们，在自发的反应中自由能总是下降的（ΔG 是负值）。然

而，当反应达到平衡时，自由能将不再变化（$\Delta G=0$）。密闭体系的平衡态是自由能最低的状态。

在反应趋近平衡态的过程中失去的自由能，或者转化为热，或者使熵值增加。在此，我们不会试图定义熵，只能说熵是衡量混乱度的标准。混乱度越高，熵值越大。熵的存在意味着许多自发的化学反应（伴随有自由能的净下降）并不需要提供激发热能。例如，当氯化钠（NaCl）溶于水时，体系吸热而不是放热。虽然如此，自由能仍然净下降，因为钠离子和氯离子从固体溶入水的过程增加了体系的混乱度。

K_{eq}与 ΔG 成指数关系

明显地，化学键越强，相应地伴随其所释放的自由能（ΔG）越多，以成键状态存在的原子的比率越大。这一常识以物理化学方程定量表示如下：

$$\Delta G=-RT\ln K_{eq} \quad 或 \quad K_{eq}=e^{-\Delta G/RT} \qquad [方程 3\text{-}4]$$

在这里 R 是气体常数，T 是绝对温度，ln 是以 e 为底数的对数（以 K_{eq} 为幂），K_{eq}是平衡常数，e=2.718。

将 R 值（1.987cal/deg · mol）和 T（当温度为 25℃时等于 298）代入公式中告诉我们，如果所有的反应物以摩尔浓度表示，低至 2kcal/mol 的 ΔG 能够使一个成键反应进行完全（表 3-1）。

表 3-1　25℃时平衡常数与 ΔG 的数字关系

平衡常数（K_{eq}）	自由能 ΔG（kcal/mol）
0.001	4.089
0.01	2.726
0.1	1.363
1.0	0
10.0	−1.363
100.0	−2.726
1000.0	−4.089

共价键是很强的

伴随自由原子，如氢或氧，共价键形成的 ΔG 以符号表示是极大的负值，通常有 −50～−110kcal/mol。方程 3-4 告诉我们，成键反应的 K_{eq}会相当大，所以保持未成键状态的氢原子或氧原子的比率将会极低。如果我们有 1mol/L 的反应原子，且 ΔG 为 −100kcal/mol，当反应达到平衡时每 10^{40} 个原子中仅会有一个保持未成键状态。

生物体系中的弱化学键

范德华力、疏水键、氢键和离子键是生物体系中几种重要的弱化学键。我们将要看到，有时氢键和离子键之间的区别并不那么严格。

弱化学键具有 1～7kcal/mol 的能量

范德华力是最弱的化学键。它的能量（1～2kcal/mol）仅仅比热运动的动能稍大。氢键和离子键的能量介于 3～7kcal/mol。

在溶液体系中，几乎所有分子都与其邻近原子建立一些弱化学键。任何分子都有形成范德华力的能力，而氢键或离子键只能在带有净电荷（离子）或电荷分布不平均的分子之间形成。某些分子拥有形成多种弱化学键的能力。然而考虑到能量因素，我们知道分子总是更倾向于形成较强的化学键。

生理温度下弱化学键不断形成和断裂

最强的弱化学键的能量仅是 25℃时（热）运动的平均动能（0.6kcal/mol）的 10 倍左右。由于热运动的动能具有明显的扩散性，在生理温度下具有许多这样的分子，它们拥有足够的动能来打断最强的弱化学键。

极性与非极性分子的区别

所有形式的弱相互作用都是基于电荷间的吸引力。电荷间的分离可能是永久的，也可能是暂时的，这取决于对象原子。例如，氧分子（O∶O）的两个氧原子间具有对称的电荷分布，所以两个原子都不带电荷；相反，水分子（H∶O∶H）的成键电子不是被平衡共享，因此具有不对称的电荷分布（图 3-3）。这些成键电子与氧原子更牢固的结合，导致氧原子携带一定的负电荷，而两个氢原子携带相等的正电荷，正电荷的中心在负电荷中心的另一侧。分离的正负电荷的合力被称为电荷偶极矩（dipole moment）。电子的不平均共用反映了成键原子对电子的不同吸引力。具有获取电子倾向的原子称为负电性（electronegative）原子；具有失去电子倾向的原子称为正电性（electropositive）原子。

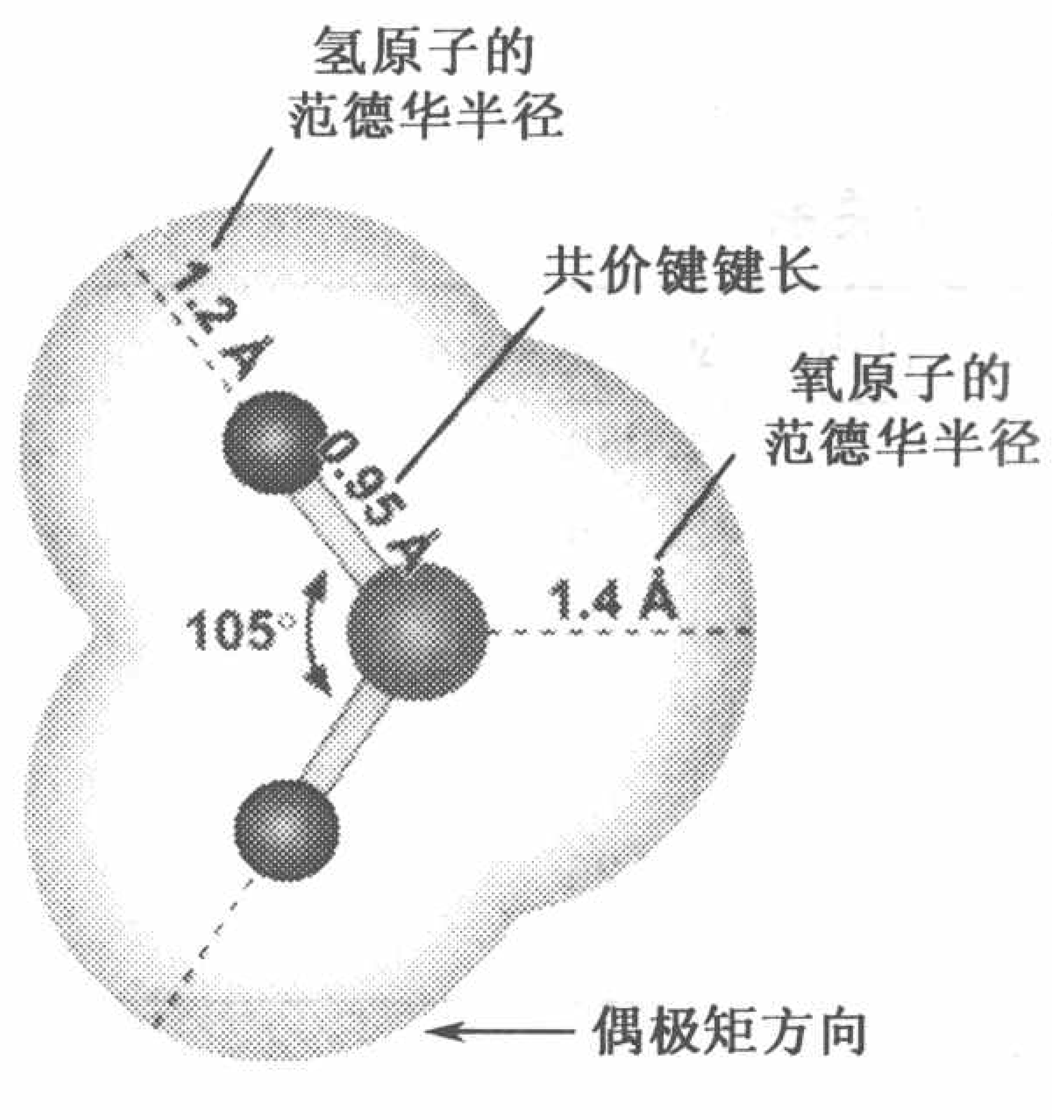


图 3-3 水分子的结构。

具有偶极矩的分子（如 H_2O）被称为极性分子（polar molecule）；没有有效偶极矩的分子称为非极性分子（nonpolar molecule）。例如，在甲烷（CH_4）中碳原子和氢原子对它们共用的电子对具有相似的吸引力，所以任何一方都不带明显电荷。

一个分子的电荷分布能够被其周围存在的其他分子所影响，当后者是极性分子时影响尤其明显。这种影响能够使非极性分子具有轻微的极性特性。如果后者没有极性，它的存在通过电荷分布波动仍然能够改变非极性分子。但是，这种削弱的影响所产生的电荷分离要远远弱于极性分子，导致极低的反应能量和弱化学键的自发生成。

范德华力

范德华力（van der Waals bonding）来源于两个相距足够近的原子之间的非特异吸引

力。它产生的基础不是永久的电荷分离的存在，而是分子的靠近所引起的电荷轻微波动。范德华力作用于所有类型的分子，不论是极性分子还是非极性分子。由于键能与键长的 6 次方成反比，因此，范德华力受相互作用的分子之间的距离的影响极大(图 3-4)。

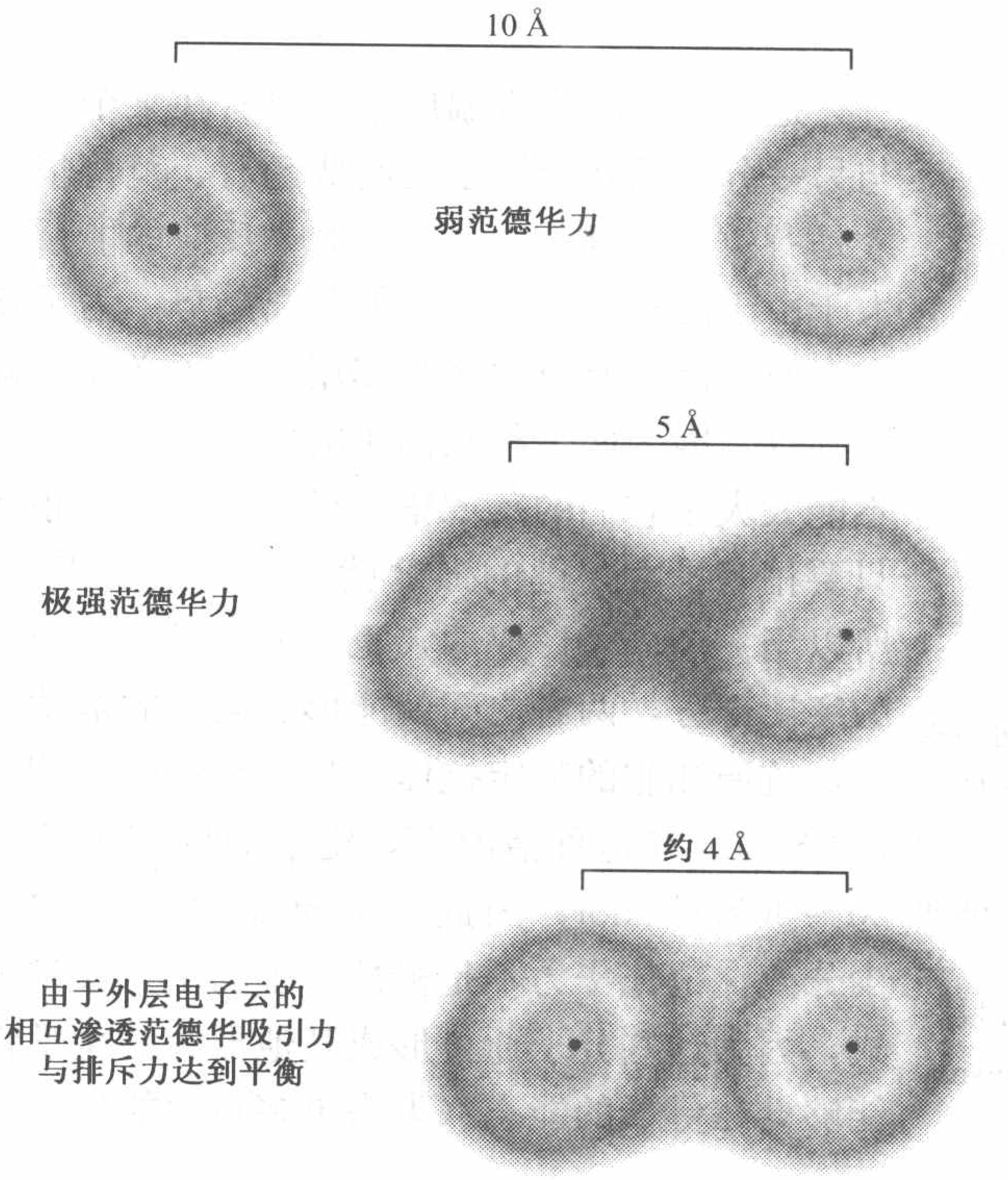


图 3-4　范德华力随距离变化的改变。本图中所示原子为稀有惰性气体氩原子。(来源：Adapted from Pauling L. 1953, General chemistry, 2nd. edition, p. 322. Copyright 1953 by W. H. Freeman. Used with permission.)

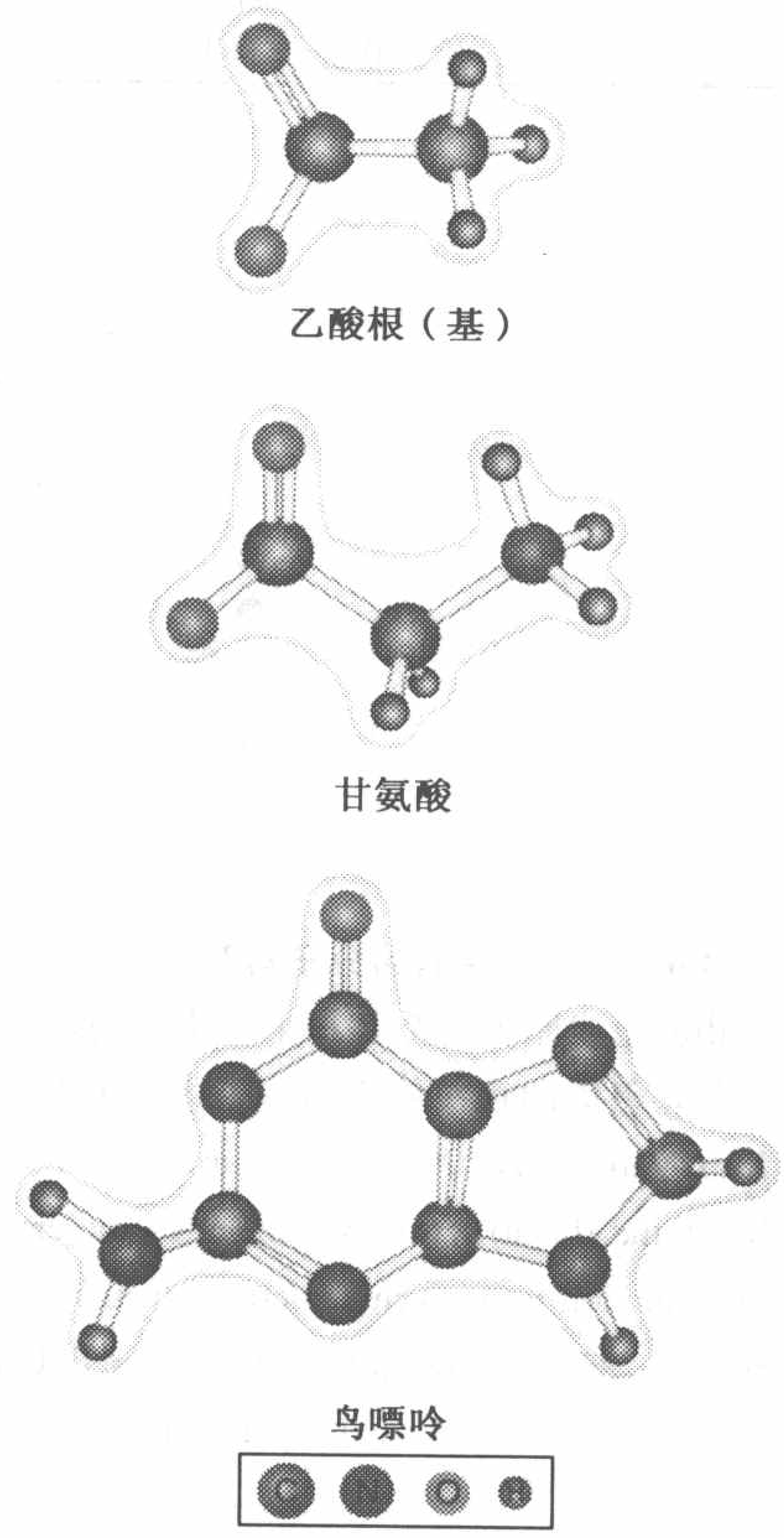


图 3-5　几种保持范德华半径的分子，原子分别以紫色、蓝色和橙色表示。

当原子间距离更短时，更强的范德华排斥力同样存在，这种排斥源于所涉及原子的外电子层的重叠。对每种原子，范德华吸引力和排斥力会在某一特定距离达到平衡，这个距离即所谓的范德华半径（van der Waals radius）（图 3-5 和表 3-2）。随着原子大小的增加，相距距离为范德华半径的两个原子间的范德华力的能量增大。对普通原子，其能量大约为 1kcal/mol，仅比室温下的分子平均热能（0.6kcal/mol）略大。

表 3-2　生物分子中原子的范德华半径

原子	范德华半径（Å）
H	1.2
N	1.5
O	1.4
P	1.9

续表

原子	范德华半径（Å）
S	1.85
CH_3 基团	2.0
芳香族分子的半厚度	1.7

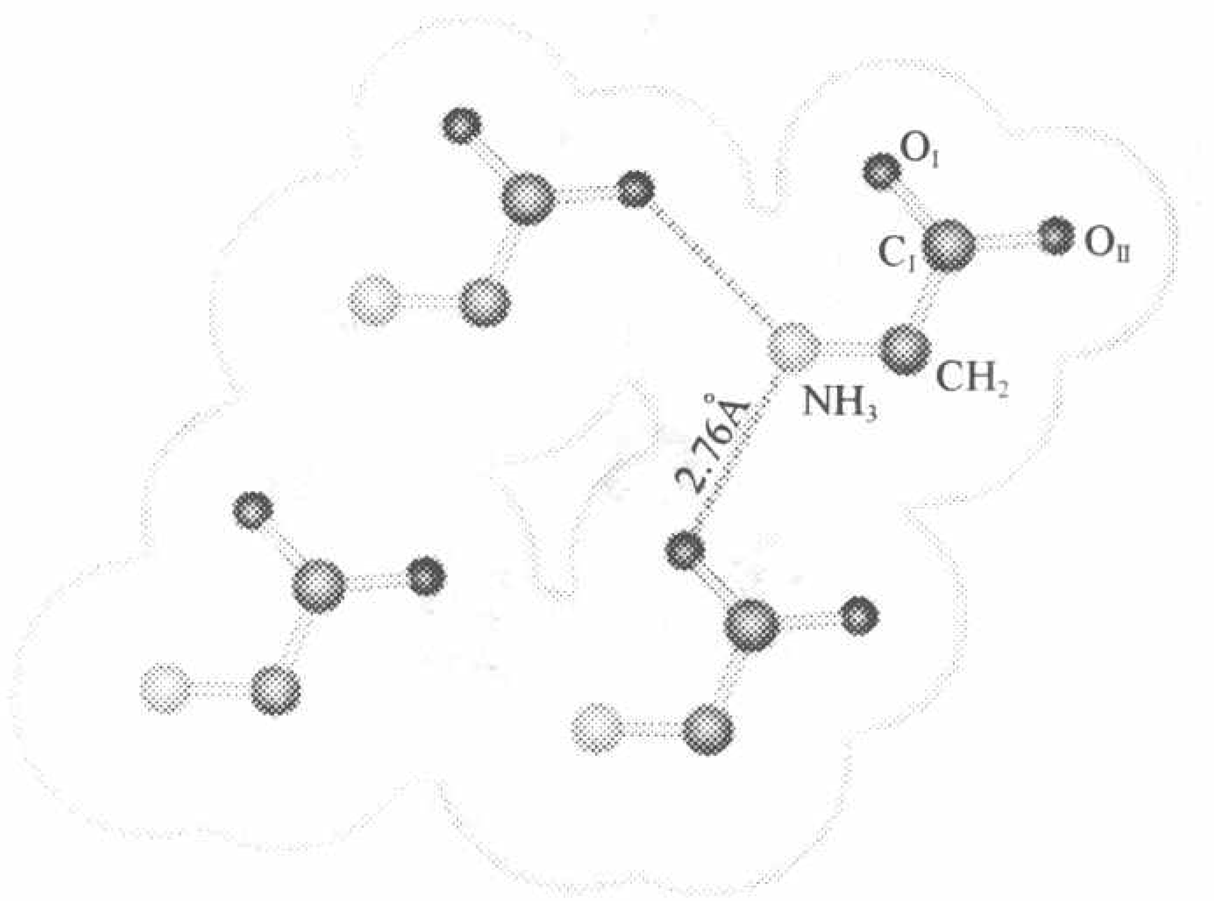


图 3-6 甘氨酸结晶的分子排列。分子的空间结构是由其组成原子的范德华半径决定的，N—HO 键例外，氢键的存在使它缩短了。（来源：Adapted from Pauling L. 1960. The nature of the chemical bond and the structure of molecules and crystals: An introduction to modern structure chemistry, 3rd edition, p. 262. Copyright© 1960 Cornell University. Used by permission of the publiser.）

这意味着在生理温度下，只有在某个分子与另一分子的几个原子之间成键时，范德华力才是有效的成键力。只有这时，反应的能量才大于由于随机热运动带来的分离趋势。为了使数个原子能够有效地相互作用，分子必须精确地吻合，因为两个相互作用的原子之间的距离不能太大于它们的范德华半径（图 3-6）。距离只要稍有扩大，键的强度就会迅速趋向于零。所以当一个分子形成的穴正好配合另一个分子的突出部分，两者之间就能形成最强的范德华力，抗原和它的特定抗体就是一个最好的例子（图 3-7）。在这种情况下，键能有时能够达到 20～30kcal/mol，因此，抗原-抗体复合物极少分离。范德华力不是极性分子的主要成键模式，因为这些分子通过形成其他形式的化学键能够处在更低的能态（失去更多自由能）。

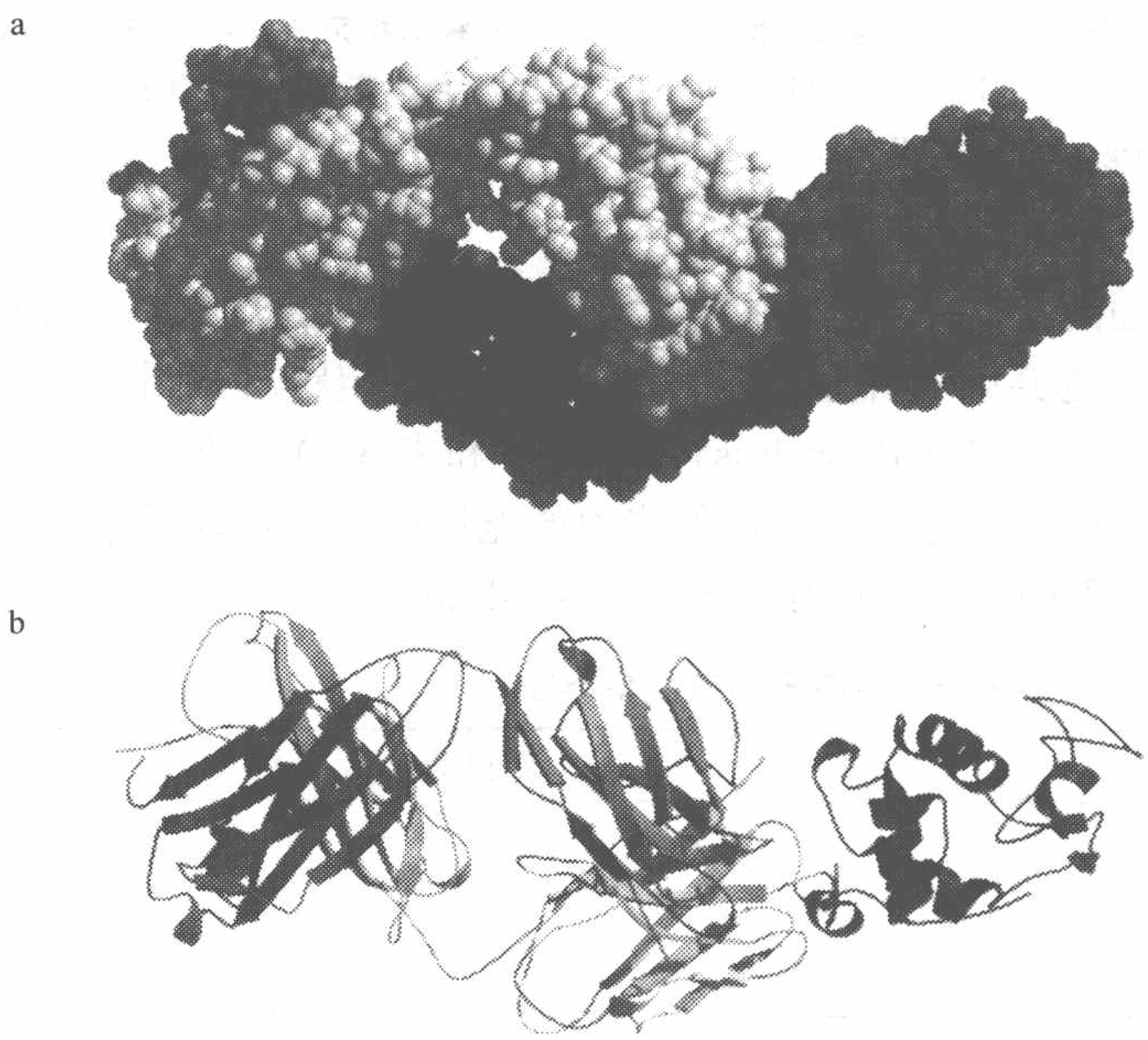


图 3-7 抗原-抗体反应。所示结构为 Fab D 1.3 和溶菌酵素所形成的复合物。（Fischmann T. O., Bentley G. A., Bhat T. N., Boulot G., Mariuzza R. A., Phillips S. E., Tello D., and Polijak R. J. 1991. J. Biol. Chem. 266: 12915.）

氢键

氢键形成于一个带部分正电的共价键供体氢原子和一个带部分负电的共价键受体原子之间（图 3-8）。例如，氨基（—NH_2）的氢原子被带负电的酮基（—C═O）的氧原子吸引。有时，形成氢键的原子属于某一带一个单位电荷的基团（如 NH_3^+ 或 COO^-）；有时，供体氢原子和负电性受体原子都带不足一个单位电荷。

与氧原子（O—H）或氮原子（N—H）共价连接的氢原子形成了生物学上最重要的氢键。而且，氧原子和氮原子也是常见的负电性受体原子。表 3-3 列出了一些最重要的氢键。如果不考虑环境中的水分子，氢键键能介于 3～7kcal/mol，受体原子与供体原子间的电荷差越大键越强。氢键虽然弱于共价键，却明显强于范德华力。因此，氢键将使两个原子间的距离小于范德华半径，但不会小于由共价键连接的两个原子间的距离。

表 3-3　生物学中重要氢键的大约键长

键	氢键近似长（Å）
O —H……O	2.70±0.10
O —H……O^-	2.63±0.10
O —H……N	2.88±0.13
N —H……O	3.04±0.13
N^+—H……O	2.93±0.10
N —H……N	3.10±0.13

与范德华力不同的是，氢键有高度的方向性。在最强的氢键中，氢原子明显的朝向受体原子（图 3-9）。如果氢原子的偏向超过 30°，键能就比较低了。而且由于氢键的形成需要分子中存在互补的供体氢原子和受体基团，氢键比范德华力更具特异性。

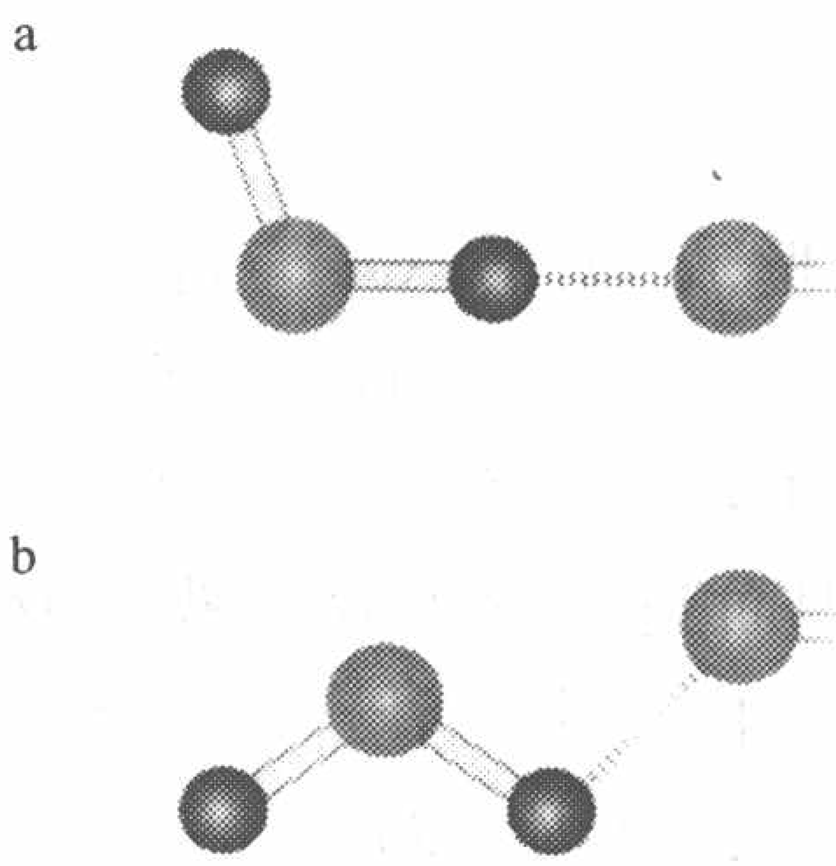


图 3-9　氢键的方向性。 a. 共价键O—H的矢量直接指向受体氧原子，形成强化学键。b. 矢量没有指向氧原子，形成弱化学键。

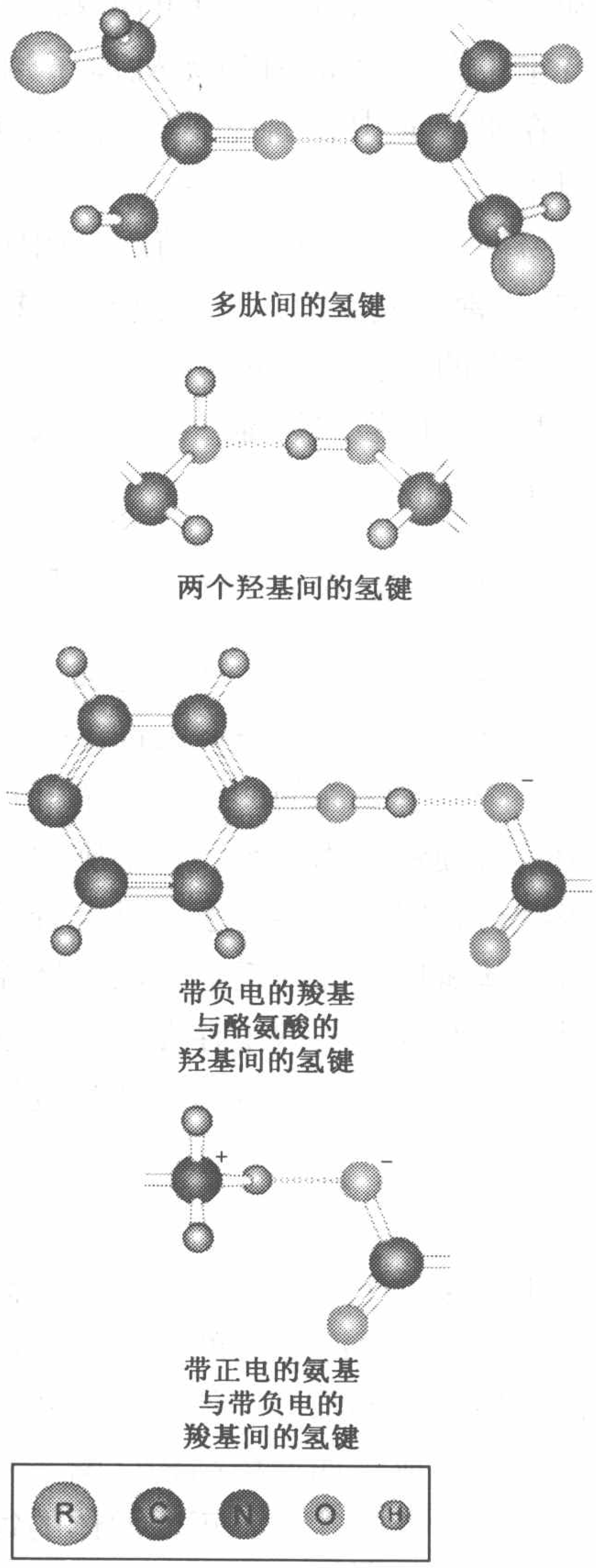


图 3-8　生物分子中的氢键举例。

某些离子键也是氢键

许多有机分子由于带有一个或多个净正电荷或负电荷而具有了离子特性。例如，带有负电的单核苷酸含有带负电的磷酸基团，而所有氨基酸（脯氨酸除外）都带有负电性的羧基（COO^-）和正电性的氨基（NH_3^+），各带一个单位电荷。这些带电基团的电荷通常被附近带相反电荷的基团所中和。带相反电荷的基团间的静电力称为离子键。在水溶液中，离子键的平均键能一般为 5kcal/mol。

在一般情况下，无机阳离子（如 Na^+、K^+ 或 Mg^{2+}）或阴离子（如 Cl^- 或 SO_4^{2-}）中和了离子化的有机分子的电荷。在水溶液中，中和电荷的阳离子和阴离子并不占据特定位置，这是因为无机离子通常带有水化层，因而并不与带电基团直接接触。所以在水溶液中，围绕无机阳离子或阴离子的静电力并不是决定有机分子外形的重要因素。

另一方面，带相反电荷的基团所形成的氢键将具有高度的方向性。例如，COO^- 和 NH_3^+ 经常形成氢键。由于这种键要强于由带不到一个单位电荷的基团所形成的化学键，它们的键长也会相应的短些。而带一个单位电荷的基团也可以和带不到一个单位电荷的基团形成强的氢键。例如，氨基（NH_2）的氢原子能与羧基（COO^-）的氧原子形成强化学键。

弱相互作用需要互补的分子表面

只有当相互作用表面非常邻近时，弱化学键才能起作用。而只有当分子表面具有互补结构（complementary structure）时，也就是说一个分子表面的突出部分（或正电）与另一个分子表面的凹穴（或负电）相称时，这种邻近才有可能实现，即相互作用分子必须具有锁钥结构关系。在细胞中，这一要求通常意味着一些分子不能和同类型的另一分子结合，因为这些分子不具有分子内相互作用所必需的对称性。例如，一些极性分子拥有供体氢原子却缺乏适合的受体原子，而另一些分子虽然能够接受氢键却没有氢原子来供给。此外，细胞中存在着许多这样的分子，由于拥有必要的对称性，它们能够形成强分子内相互作用，水分子就是这样一个重要的例子。

水分子形成氢键

生理条件下，水分子几乎从不离子化形成 H^+ 和 OH^- 离子。取而代之的是，它们以极性 H—O—H 分子的形式存在，其中的氢原子和氧原子都形成了强氢键。每个水分子的氧原子能够结合其他分子的两个氢原子，而每个氢原子可以结合一个邻近氧原子。这些氢键构成正四面体结构（图 3-10），因而当水分子处于液态或固态时，每个水分子总有 4 个最近的邻居，正四面体的每一个方向一个。处于冰的形态的水分子与邻近水分子的结合将会非常牢固，邻近水分子的数目也是固定的。如果温度大于冰点（0℃），热运动的能量就能打断氢键，允许水分子不断变更其最邻近的水分子。然而，即使是液态水，大多数水分子在任一瞬间内还是由 4 个强氢键连接着。

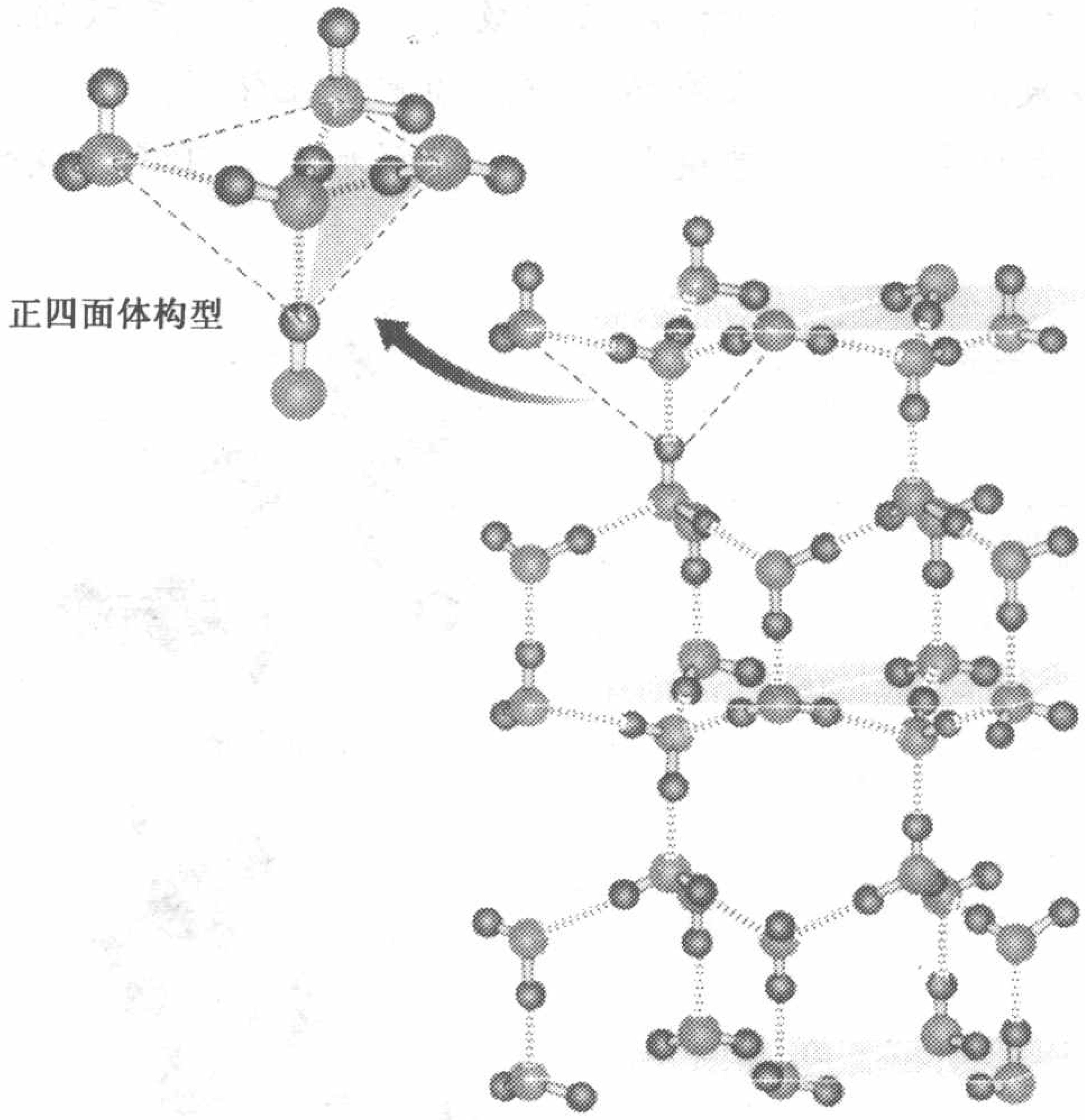


图 3-10　水分子所形成的晶格。 水分子间形成氢键所获得的能量促使其形成正四面体结构。氧原子以大圆表示，氢原子以小圆表示。尽管分子的温度决定了排列的严格性，如图所示的结构不但在冰中，在水中也还是占优势。（来源：Adapted from Pauling L. 1960. The nature of the chemical bond and the structure of molecules and crystals: An introduction to modern structure chemistry, 3rd edition, p. 262. Copyright© 1960 Cornell University. Used by permission of the publisher.）

水溶液中分子间的弱化学键

次要化学键的平均能量尽管低于共价键，却毫无疑问地大于热能，因而能够确保大部分水溶液中的分子与其他分子形成次要化学键。方程 3-4 经修正适合于液体中高浓度分子的情况下，可以给出分子的成键比率。由此我们可以知道，在生理温度下低至 2～3kcal/mol 的反应能都足以使大多数分子形成最大量的强次要化学键。

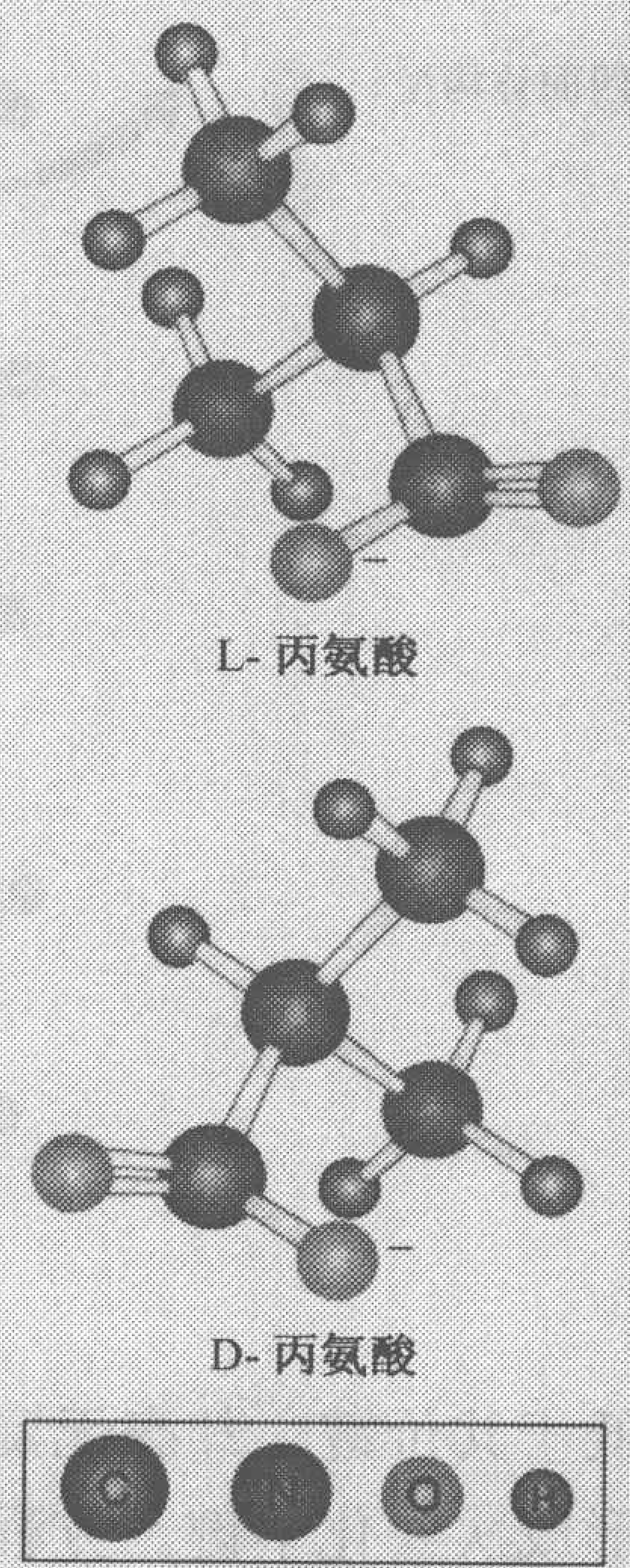
溶剂分子在特定瞬间内会显著地影响溶质分子的特定结构，这不仅仅是因为分子都具有特定外形，而且还因为不同的分子能够形成不同的次要化学键。所以，一个分子将会不停运动，直到它与一个能与之形成最强化学键的分子相邻。

毫无疑问，溶液是非静态的。由于热运动的扰乱，溶液的特定结构不断从一个状态变化到另一具有同等内能的状态。在生物体系中同等重要的一个事实是，新陈代谢持续地将一个分子转变为另一个分子，因此自动地改变了所形成的次要化学键。细胞的溶液态结构不但被热运动不停扰动，而且受其溶质分子在新陈代谢过程中变化的影响。

框 3-1　分子外形的唯一性和选择性结合的概念

即使大多数细胞内的分子是由少数化学基团（如 OH、NH_2 和 CH_3）所构成，但分子对其邻近分子的选择仍然有着很大的特异性。这是由于每个分子都有其唯一的成键性质。一个非常明显的证明来自于立体异构物的特异性。例如，蛋白质总是由 L-氨基酸构成，而不是其镜像 D-氨基酸（框 3-1 图 1）。尽管 L-氨基酸和 D-氨基酸具有完全相同的共价键，它们和不对称分子的结合特性却十分不同。因此，大部分酶对 L-氨基酸特异。如果一种特异性的酶能够和 L-氨基酸结合，那么它就不能结合 D-氨基酸。

细胞内大多数分子仅能与少数其他分子形成有效的弱化学键，部分由于生物系统中的大多数分子存在于溶液环境中。细胞中化学键的形成不仅依赖于两个分子是否连接紧密，而且依赖于此化学键是否比与溶剂水分子可能形成的其他化学键有利。



框 3-1 图 1　丙氨酸的两种立体异构体。（来源：Adapted from Pauling L. 1960. The nature of the chemical bond and the structure of molecules and crystals: An introduction to modern structure chemistry, 3rd edition, p. 465. Copyright© 1960 Cornell University. Used by permission of the publisher. And from Pauling L. 1953, General chemistry, 2nd edition, p. 498. Copyright 1953 by W. H. Freeman. Used with permission.）

倾向于形成氢键的有机分子是水溶性的

每原子基团的氢键能量是远大于其范德华力的，所以与范德华力相比，分子更容易形成氢键。例如，如果我们试图在水中溶解不能形成氢键的化合物，如苯，水分子和苯分子会迅速分开，水分子间形成氢键而苯分子间以范德华力聚集。因此在水中溶解非氢键有机分子是不可能的。

另一方面，那些含有大量能够形成氢键的基团（如═O 或 OH）的极性分子（如葡萄糖和丙酮酸）能够溶解于水（即它们是亲水的，而不是疏水的）。当这类基团对水分子晶格的插入打断水-水氢键时，立即导致了极性有机分子和水分子间氢键的形成。然而在能量上，这种变化的排列一般不如水分子之间的排列稳定，因此即使是极性最大的分子一般也有一定的溶解度（框 3-1）。

所以，几乎所有细胞需要的分子，不管是通过食物还是生物合成获取，在水中都或多或少是不溶的。通过热运动，这些分子偶尔与其他分子发生碰撞，直到它们与互补的

分子表面结合，释放水分子以参与水相反应。

疏水“键”稳定大分子

水分子排斥非极性分子的强烈趋势经常被称作疏水键（hydrophobic bonding）。一些化学家喜欢称水溶液中所有非极性分子间的化学键为疏水键（图 3-11）。从某种意义来说这种叫法是错误的，因为它试图强调的现象是化学键的不存在，而不是存在。（非极性分子间易于形成的化学键是基于范德华吸引力。）另一方面，疏水键这一叫法又很有用，因为它强调了这样一个事实，即非极性分子会试图排列在一起以避免与水分子接触。不论是对蛋白质分子之间和蛋白质与其他分子的复合物的稳定性，还是对蛋白质在细胞膜上的分布，疏水作用都非常重要，它可能占据了蛋白质折叠时总自由能的一半。

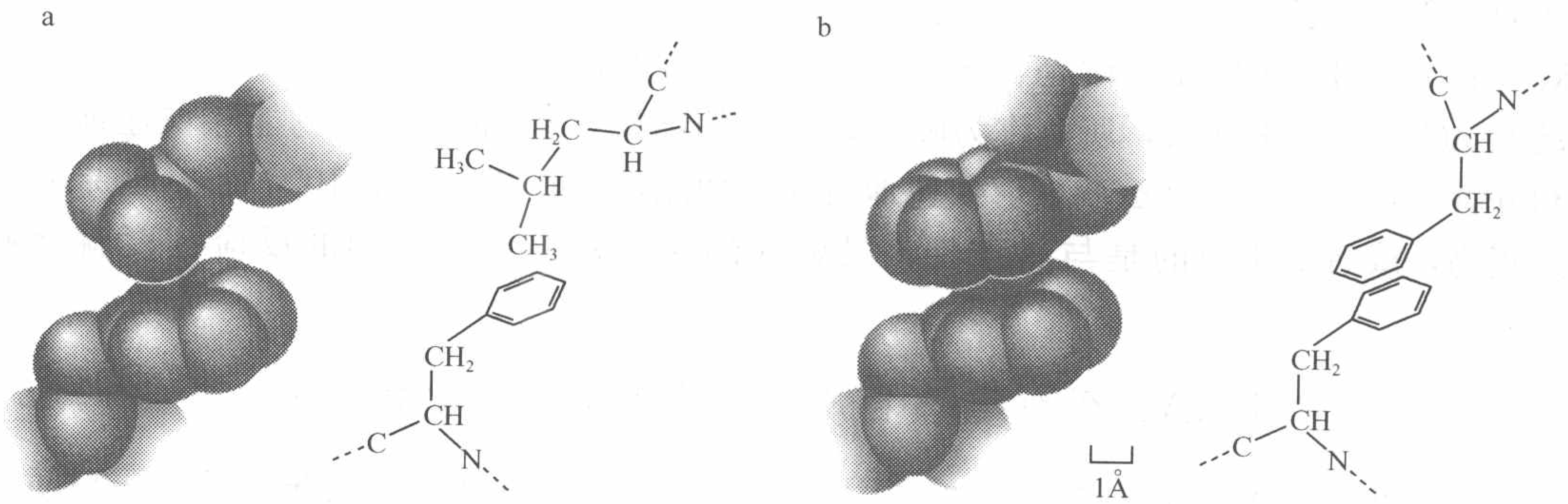


图 3-11　氨基酸非极性侧链间的范德华力（疏水性）。氢原子没有单独标出，为了图示的清晰，范德华半径缩小了 20%，实心图旁的结构式表示原子的排列情况。a. 苯丙氨酸-亮氨酸键；b. 苯丙氨酸-苯丙氨酸键。（来源：Adapted from Scheraga H. A.，The proteins，2nd edition，p. 527. Copyright© Harold Scheraga. Used with permission.）

例如，考虑丙氨酸和甘氨酸在水溶液中与一个拥有与丙氨酸互补表面的分子成键时释放能量的差异。丙氨酸中有一个甲基，而甘氨酸却没有。当丙氨酸与这个分子成键时，甲基周围的范德华力释放出 1kcal/mol 的能量，而甘氨酸成键时却不会。从方程 3-4我们可以知道，单独的如此小的能量差异仅仅会在丙氨酸和甘氨酸的成键中造成一个等于 6 的因子。然而，如此计算并没有将这样一个事实考虑在内，即水对丙氨酸的排斥度远大于甘氨酸。丙氨酸中 CH_3 的存在比甘氨酸中的氢原子更严重的扰乱了水分子的晶格。目前，由疏水性侧链扰乱水分子晶格而必须引入的修正因子还难以预测。但很可能，水倾向于排斥丙氨酸，并将它推向第三分子，这是通过比水排斥甘氨酸的力约大 2～3kcal/mol的疏水作用完成的。

由此我们得出了一个重要结论，即使最为相近的分子与第三分子成键时的能量差别（当相似分子间的差别为一个非极性基团），在细胞内溶液中比在非水溶液环境中也至少大 2～3kcal/mol。有时，能量差异可达 3～4kcal/mol，这是因为涉及的分子通常含有能形成氢键的极性基团。

ΔG 具有的 2～5kcal/mol 的优势

我们已经知道，一个次要化学键的能量（2～5kcal/mol）能够有效确保一个分子与

另一分子的特定基团优先成键。而且，这种能量差不多会大到在细胞中形成严格的晶格排列；也就是说，细胞内不会结晶，如果次要化学键的能量大上数倍，这种情况就会发生。更大的能量差意味着次要化学键极少断裂，这将会导致与细胞生存不相容的低溶质扩散率。

弱化学键使酶与底物结合

次要作用力是酶与其底物初始结合的必需基础。酶并非任意地与所有分子结合，而是仅与其底物有显著的亲和性。

因为酶对化学反应的两个方向都有催化作用，所以它们必须对反应分子的两种形式都有特异的亲和性。在某些情况下，可以测量酶与其底物结合的平衡常数（方程 3-4），从而计算出成键的 ΔG。由此，可以得到关于成键类型的提示。一些 ΔG 在 5～10 kcal/mol间的强次要化学键是特异性酶-底物反应的基础。另一个值得注意的是，化学键的 ΔG 极少异常地大，所以作为随机热运动的结果，酶-底物复合物能够快速地形成和分解。这解释了为什么酶能够快速地行使其功能，有时甚至快至 10^6 次/s。如果酶与其底物，或者更重要的是与其产物，以更强的化学键结合，它们的反应速度就会慢得多。

大多数蛋白质-DNA、蛋白质-蛋白质相互作用由弱化学键介导

正如我们在本书中将要看到的，蛋白质与 DNA 的相互作用以及蛋白质与蛋白质的相互作用构成了细胞生命活动的核心，这些生命活动包括如何识别信号并做出反应、基因的表达、DNA 的复制、修复和重组等，当然还包括这一系列过程的调控。同样地，这些相互作用是通过在本章中所描述的这类弱化学键来实现的。尽管单个化学键的能量很低，但是任意两个相互作用分子间的许多这类化学键的共同作用保证了反应的亲和性以及特异性。

在第 5 章，我们将要具体讨论蛋白质是如何构建的，它们是怎样形成特定的结构，以及它们是如何与 DNA 或蛋白质结合的。

小结

细胞中许多重要的化学过程并不包含共价键的形成和断裂。大多数分子的细胞定位依赖于弱（或者说次要）的吸引力或排斥力；而且，弱化学键在决定许多分子（尤其是大分子）外形上起着重要作用。最重要的几种弱化学键是氢键、范德华力、疏水作用和离子键。即使这类键相对较弱，它们还是足够强，以保证正确的分子（或原子团）相互作用。例如，酶分子表面的外形唯一适合于与其底物分子的结合。

不论是弱化学键还是强共价键，所有化学键的形成遵循热力学定律。如果反应的结果是释放自由能（负 ΔG），化学键就能形成。为了打断化学键，必须给予相同大小的自由能。由于在原子间形成共价键会导致很大的负 ΔG，共价连接的原子几乎不会自发

分开。相反地，伴随弱化学键形成的 ΔG 仅仅是生理温度下分子平均热运动能量的数倍。在活细胞中单个弱化学键频繁的形成和断裂。

具有极性（带电）基团的分子拥有与非极性分子（其中的电荷平均分布）完全不同的相互作用方式。极性分子能够形成氢键，而非极性分子仅能形成范德华力。最重要的极性分子是水分子，每个水分子可与其他水分子形成 4 个氢键。极性分子大多是水溶性的（程度不同），而非极性分子不能溶于水，这是因为它们不能与水分子形成氢键。

任何一个特定分子具有其唯一的分子外形，这严格限制了它能与之形成强次要化学键的分子的数量。强次要相互作用不但需要两个分子作用表面间的互补关系（锁钥关系），还需要许多原子的参与。尽管只由 1～2 个次要化学键连接的分子经常发生解离，但是这些弱化学键的综合能够使分子形成稳定的聚集体。DNA 双螺旋结构从不自发解离的事实说明，在这样的聚集体里，极端的稳定是可能的。

（时亮　译　吴东颖　刘国振　刘韧　校）

参考文献

Branden C. and Tooze J. 1999. *Introduction to protein structure*, 2nd ed. Garland Publishing, New York.

Creighton T.E. 1992. *Proteins: Structure and molecular properties*, 2nd ed. W.H. Freeman, New York.

——— 1983. *Proteins*. W.H. Freeman, San Francisco.

Donohue J. 1968. Selected topics in hydrogen bonding. In *Structural chemistry and molecular biology* (ed. A. Rich and N. Davidson), pp. 443–465. W.H. Freeman, San Francisco.

Fersht A. 1999. *Structure and mechanism in protein science: A guide to enzyme catalysis and protein folding*. W.H. Freeman, New York.

Gray H.B. 1964. *Electrons and chemical bonding*. Benjamin Cummings, Menlo Park, California.

Klotz I.M. 1967. *Energy changes in biochemical reactions*. Academic Press, New York.

Kyte J. 1995. *Mechanism in protein chemistry*. Garland Publishing, New York.

——— 1995. *Structure in protein chemistry*. Garland Publishing, New York.

Lehninger A.L. 1971. *Bioenergetics*, 3rd ed. Benjamin Cummings, Menlo Park, California.

Lesk A. 2000. *Introduction to protein architecture: The structural biology of proteins*. Oxford University Press, New York.

Marsh R.E. 1968. Some comments on hydrogen bonding in purine and pyrimidine bases. In *Structural chemistry and molecular biology* (ed. A. Rich and N. Davidson), pp. 485–489. W.H. Freeman, San Francisco.

Morowitz H.J. 1970. *Entropy for biologists*. Academic Press, New York.

Pauling L. 1960. *The nature of the chemical bond*, 3rd ed. Cornell University Press, Ithaca, New York.

Tinoco I. (ed.), Sauer K., Wang J.C., and Puglisi J.D. 2001. *Physical chemistry: Principles and applications in life sciences*, 4th ed. Prentice Hall College Division, Upper Saddle River, New Jersey.

第4章 高能键的重要性

本章概要

•

供能分子是热不稳定的

•

酶在生物反应中降低反应的活化能

•

生物分子中的自由能

•

生物合成反应中的高能键

•

基团转移反应中前体的活化

前一章从热力学的观点讨论了弱化学键的形成。每当一个潜在的弱化学键被提及，我们都要问一个问题：它的形成导致了自由能的获取还是释放？只有当 ΔG 是负值时热力学平衡才有利于该反应。该原则同样适用于共价键。酶经常参与共价键的形成和断裂，但这一事实并不能改变对负 ΔG 的需要。

然而简单举例来说，细胞中许多重要共价键的形成似乎违背了热力学定律，尤其是那些将小分子结合成高分子的化学键。这类键形成的过程是自由能增加的过程。最初有人认为，细胞具有特殊的违反热力学定律的活动能力，这就是“生命的秘密”所在。

但是现在人们已经明白，这些生物合成过程并不违背热力学定律，而是基于那些与初始假设不同的反应。例如，浓缩的核苷酸不能生成核酸；葡萄糖残基不能直接生成糖原；氨基酸单元不能生成蛋白质。相反，单个前体首先运用储存于 ATP 中的能量转化为高能“活化”前体，然后再自发地（在特定酶的协助下）联合起来组成大分子。我们将在本章中通过举例说明这些问题，重点关注多肽（蛋白质）和核苷酸（核酸）化学键的热力学。但是，我们首先必须对共价键的一些热力学特性做简要说明。

供能分子是热不稳定的

不同分子所拥有的自由能的大小变化范围很大，这是因为共价键并不总是具有相同的能量。举例来说，氢氧原子间的共价键比氢氢原子间的或氧氧原子间的共价键明显要强，消耗 O—O 或 H—H 来形成 O—H 将会释放能量。因此，能量观点告诉我们充分浓缩的氧气和氢气的混合物将会转化为水。

一个分子以弱共价键连接时将会比以强共价键连接时具有更多的自由能。初看起来这个观点好像很矛盾，因为这意味着键越强，释放的能量越少。但是当我们了解到一个

形成强化学键的原子在此过程中已经失去大量自由能的时候，就自然能得出以上结论。因此，那些含有弱共价键并因而热不稳定的分子是最好的食物分子（供能分子）。

例如，葡萄糖就是一个理想的食物分子，因为当它被氧化为 CO_2 和 H_2O 时自由能显著降低。另一方面，CO_2 的碳氧原子间有着强共价双键，即羰基双键（carbonyl bond），不是动物的食物分子。在缺乏能量供体 ATP 时，即使有特定酶的帮助，CO_2 也不能自发转化为更复杂的有机分子。CO_2 能被植物作为主要碳源是因为光量子提供的能量在磷酸化的过程中导致了 ATP 的生成。

在生理温度下如果没有催化作用，将分子转化为其他含较低自由能分子的化学反应几乎没有速度。这是因为，即使是弱共价键相对来说也很强，在细胞中极少为热运动所打断。为了在缺少催化作用的条件下打断共价键，我们必须提供将成键原子分开的能量。当原子被分开后，它们可以与新的对象结合形成更强的化学键，在结合过程中，释放的能量等于打断旧化学键而消耗的能量与新旧键间的能量差之和（图 4-1）。

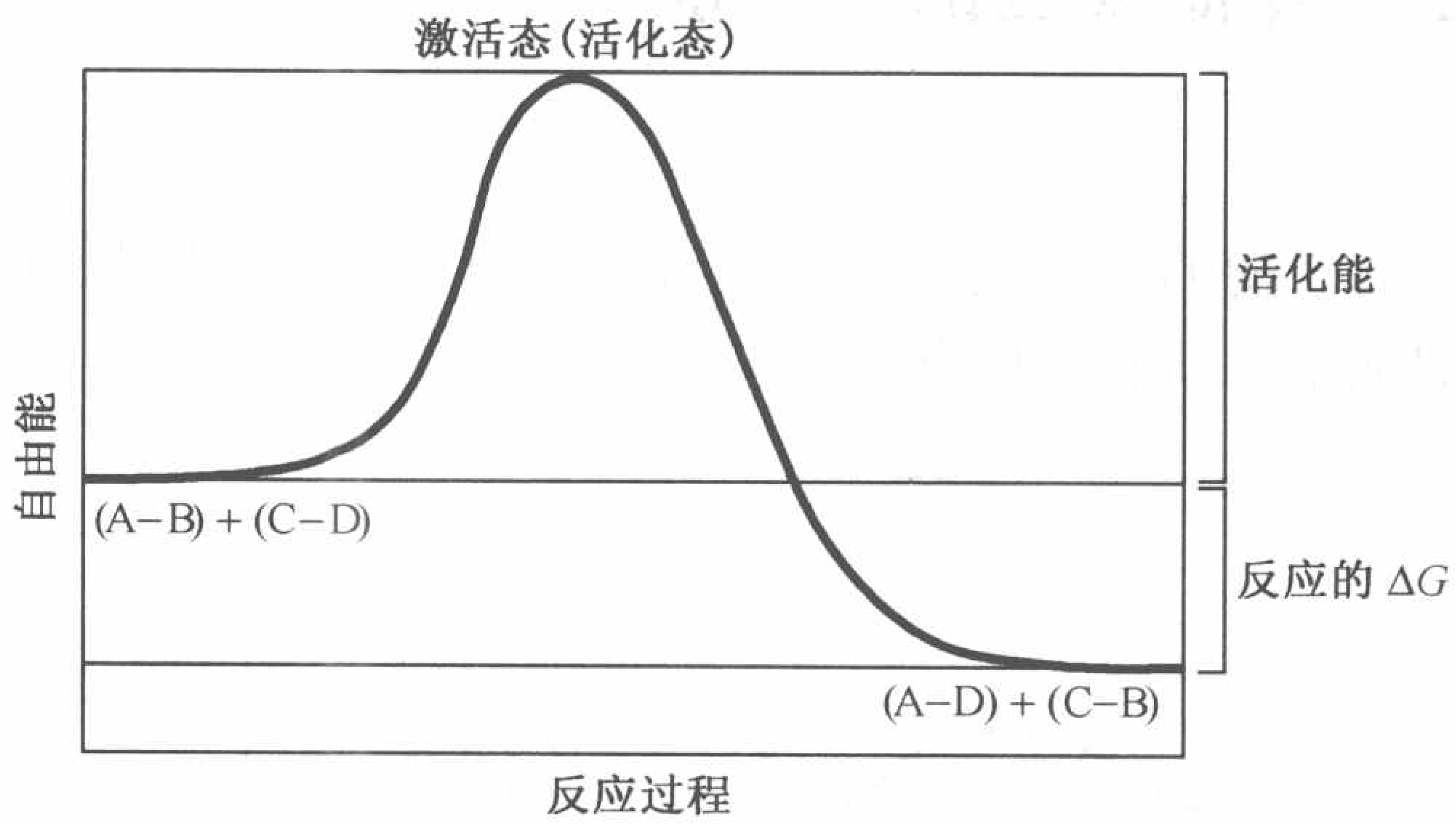


图 4-1　化学反应的活化能：(A—B)+(C—D)→(A—D)+(C—B)。这个反应伴随有自由能的下降。

在分子转化过程中必须提供的打断旧化学键的能量称为活化能（activation energy）。活化能通常小于化学键的初始能，因为分子重排一般不会产生完全自由的原子，取而代之的，两个反应分子间的碰撞是必需的，随之形成暂时的分子复合物形式，称为活化态（activated state）。在活化态，两个分子的紧密邻近使各自的化学键更加不稳定，所以打断此时的化学键就比打断自由分子的化学键所需的能量少。

细胞内大多数共价键的化学反应表示如下：

$$(A-B)+(C-D)\rightarrow(A-D)+(C-B) \quad [方程 4\text{-}1]$$

这一反应的质量方程（mass action expression）为：

$$K_{eq}=(conc^{A-D}\times conc^{C-B})/(conc^{A-B}\times conc^{C-D}) \quad [方程 4\text{-}2]$$

这里的 $conc^{A-B}$、$conc^{C-D}$ 等指的是以 mol/L 表示的反应物的浓度。同样地，在这里方程 4-3 将 K_{eq} 与 ΔG 联系起来（表 4-1）。

$$\Delta G=-RT\ln K_{eq} \quad 或 \quad K_{eq}=e^{-\Delta G/RT} \quad [方程 4\text{-}3]$$

由于反应的能量通常为 20～30kcal/mol，生理温度下活化态几乎不存在。高的活化能是阻止细胞内共价键自发重排的壁垒。这一壁垒的重要性非同寻常。没有它们，生

命将不可能出现。因为所有的原子都将处于最低能量状态，这将导致无法为后续反应储存所需的能量。但是如果没有选择性地降低某些反应的活化能的方法，生命同样不可能出现。如果想让细胞在遭受诸如电离或紫外辐射等外力造成的损害后仍能有足够快的生长速度，这种壁垒尤其是必需的。

表 4-1　K_{eq} 与 ΔG 的关系（$\Delta G=-RT\ln K_{eq}$）

平衡常数 K_{eq}	自由能 ΔG/（kcal/mol）	平衡常数 K_{eq}	自由能 ΔG/（kcal/mol）
10^{-6}	8.2	10^{-1}	1.4
10^{-5}	6.8	10^{0}	0.0
10^{-4}	5.1	10^{1}	−1.4
10^{-3}	4.1	10^{2}	−2.7
10^{-2}	2.7	10^{3}	−4.1

酶在生物反应中降低反应的活化能

生命绝对需要酶。酶的功能就是通过将分子重排的活化能降低到热运动能够提供的范围，从而加速细胞生存所必需的化学反应（图 4-2）。

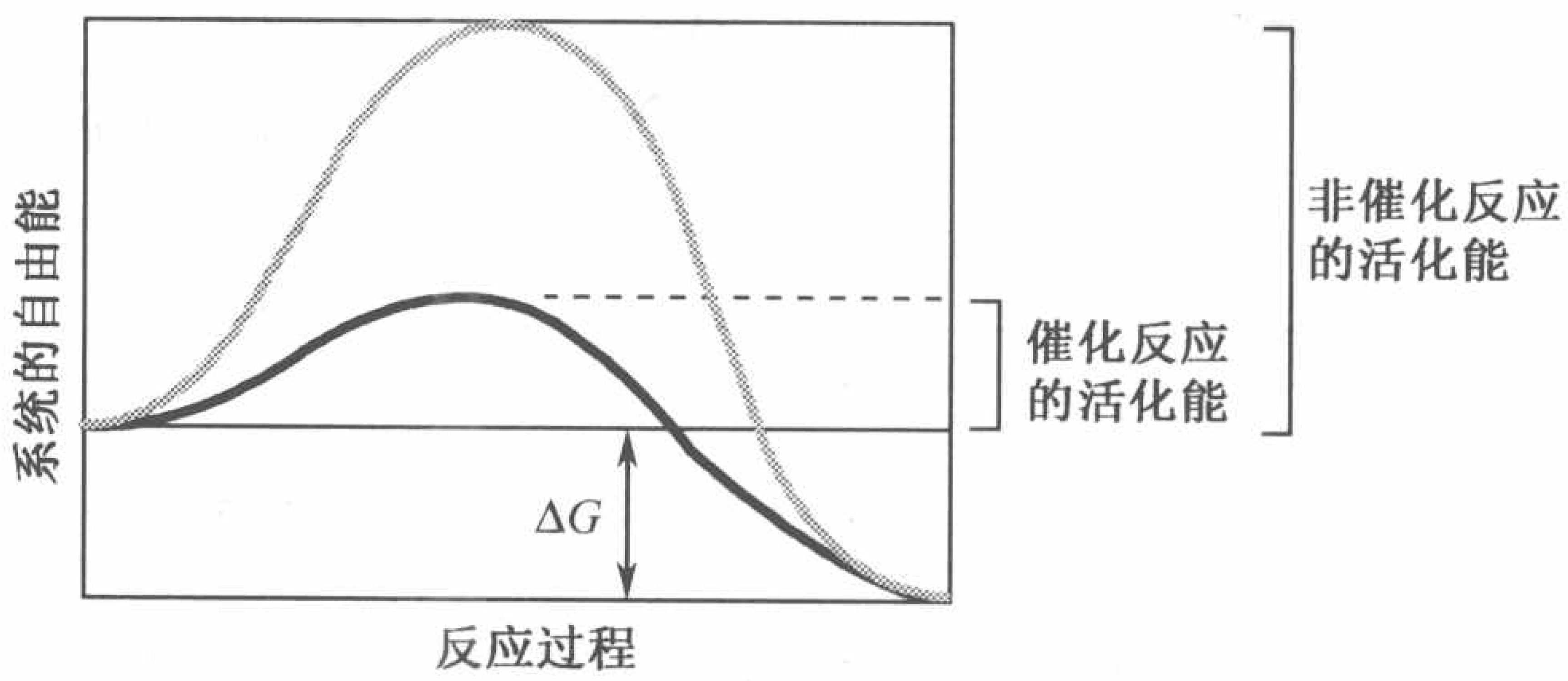


图 4-2　酶（彩色曲线）降低活化能并加快反应速度。注意因为平衡状态没有改变所以 ΔG 保持相同。

当特定的酶存在时，阻止具有最低自由能的反应物快速反应的壁垒将会消失。酶不会改变反应的平衡：它们仅仅加快了反应所达到的速度。因此，如果热力学平衡对某个分子的形成不利，酶的存在并不能提高该分子的产率。

由于酶必须催化每个细胞分子的重排，对不同分子自由能的认知并不能自然而然地告诉我们一个能量上允许的反应是否真实存在。必须考虑反应速度。只有当细胞具有适合的酶时，反应才会有意义。

生物分子中的自由能

热力学告诉我们，所有的生化路径都必须具有自由能降低的特征。降解反应路径就是一个明显的例子，在反应中伴随着热的放出，热力学不稳定的食物分子被转变为更稳

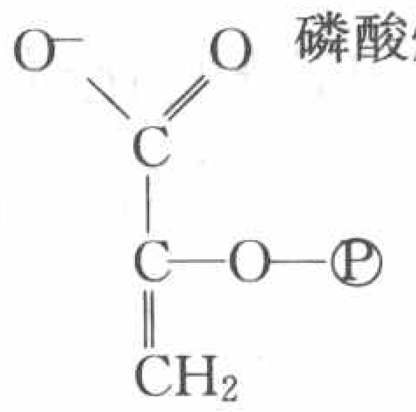
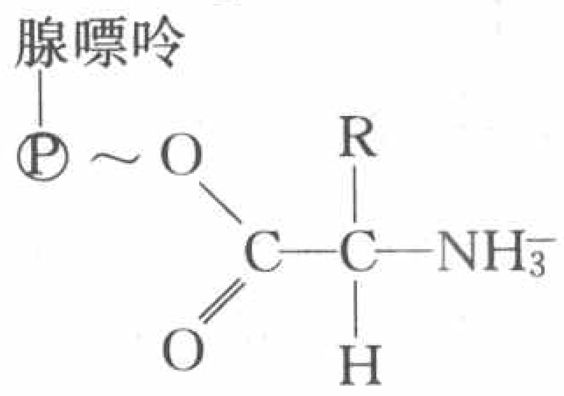
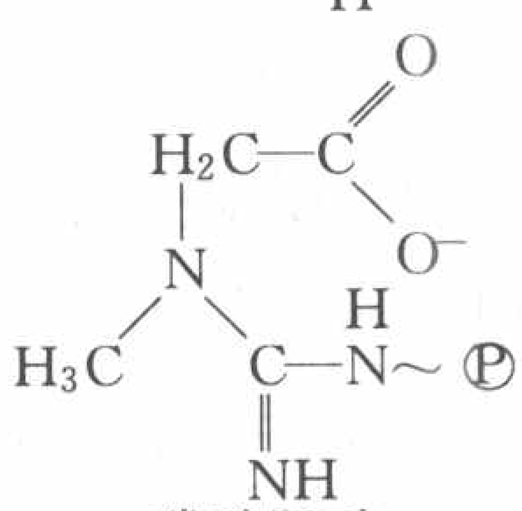
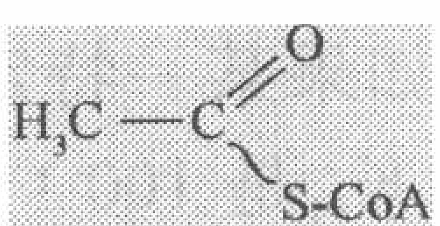
定的化合物，如 CO_2 和 H_2O。所有的降解反应路径都有两个主要目的：①为更大的有机分子的形成提供必需的作为原料的小有机分子片段；②将原始食物分子的大部分自由能以可利用的形式保存起来。通过将降解反应路径的几个步骤与储存自由能的高能分子（如 ATP）的形成相耦联，就能达到第二个目的。

食物分子的所有自由能并不是完全转化为高能分子的自由能。如果实际情况是这样，降解反应路径将不再具有自由能降低的特性，而且将不再存在有利于食物分子裂解的驱动力。相反，我们发现所有降解反应路径都会伴随着食物分子至少一半的自由能转化为热或熵。例如，据估计细胞中葡萄糖自由能的大约 40%被用来产生新的高能化合物，其余部分消散为热或熵。

具有高负值 ΔG 的高能键水解作用

一个拥有一个到数个化学键的高能分子，其化学键可被水裂解，这种作用称为水解作用(hydrolysis)，水解作用伴随有大量自由能的释放（5kcal/mol）。这种水解时导致极大负值 ΔG 的特殊的化学键称为高能键（high-energy bond）。不过这种叫法有误导作用，因为不是这些化学键的能量高而是水解时释放的自由能高。然而，高能键的叫法已被普遍接受，为了方便我们还是要继续这种用法并以符号“～”表示高能键。

表 4-2 高能键的重要类型

种类	示例分子	反应	反应的 ΔG（kcal/mol）
焦磷酸	Ⓟ～Ⓟ 焦磷酸	Ⓟ～Ⓟ ⇌ Ⓟ＋Ⓟ	$\Delta G=-6$
腺苷二磷酸	腺嘌呤—Ⓟ～Ⓟ (ADP)	ADP ⇌ AMP＋Ⓟ	$\Delta G=-6$
三磷酸腺苷	腺嘌呤—Ⓟ～Ⓟ～Ⓟ (ATP)	ATP ⇌ ADP＋Ⓟ ATP ⇌ AMP＋Ⓟ～Ⓟ	$\Delta G=-7$
磷酸化烯醇	 磷酸烯醇丙酮酸(PEP)	PEP ⇌ 丙酮酸＋Ⓟ	$\Delta G=-12$
氯酰腺苷酸		AMⓅ～AA ⇌ AMP＋AA	$\Delta G=-7$
磷酸化胍盐	 磷酸肌酸	肌氨酸～Ⓟ ⇌ 肌氨酸＋Ⓟ	$\Delta G=-8$
硫酯	 乙酰-CoA	乙酰-CoA ⇌ CoA-SH＋乙酸	$\Delta G=-8$

一个普通高能键的水解能量（7kcal/mol）要比一个葡萄糖分子一步完全降解所释放的能量（688kcal/mol）小得多。葡萄糖的一步降解是不能产生高能键的。毫无疑问，这正是生物体中葡萄糖降解需要那么多步的原因。所以，降解过程中每一步所释放的能量与高能键水解的自由能大小对应。

最重要的高能化合物是 ATP。它由无机磷酸盐Ⓟ和 ADP 所构成，形成时利用从降解反应或太阳获得的能量，后者即光合作用。然而，自然界还存在许多其他重要的高能化合物。一些直接在降解反应过程中形成，另一些则利用 ATP 的自由能形成。表 4-2 列举了数种最重要类型的高能键，每一种都含有磷或硫原子。ATP 的高能磷酸键来自于磷酸基团单元，但是，焦磷酸键（Ⓟ～Ⓟ）并非唯一一种高能磷酸键：磷酸基团与羧基氧原子的结合会产生一个高能酰键。现在我们已经清楚，在能量新陈代谢中，含有硫原子的高能键发挥着与含磷高能键几乎同等重要的作用。含有高能硫键的最重要的分子是乙酰-CoA，它是脂肪酸生物合成的主要能量来源。

高能键较大的 ΔG 分布范围（表 4-2）意味着“高能”键的叫法有些专横。一般的判断标准是该键的水解是否伴随有另一重要的生物合成反应。例如，伴随 6-磷酸葡萄糖的水解所产生的负 ΔG 是 3～4kcal/mol。但这个 ΔG 并不足以促使肽键的合成，因此这种磷脂键并不属于高能键。

生物合成反应中的高能键

由较小的分子组装形成大分子的过程通常需要自由能的输入。然而如果不伴随着自由能的净下降，生物合成反应路径（如降解反应路径）将不会发生。这意味着许多生物合成反应路径需要外在的自由能来源，而高能化合物就是这些自由能来源。生物合成中许多化学键的形成是与高能键的断裂相耦联的，因而自由能的净变化还是负值。所以细胞内高能键的寿命通常极短，它们几乎刚在降解反应中形成就被酶催化分解，以产生驱动另一反应进行完全的能量。

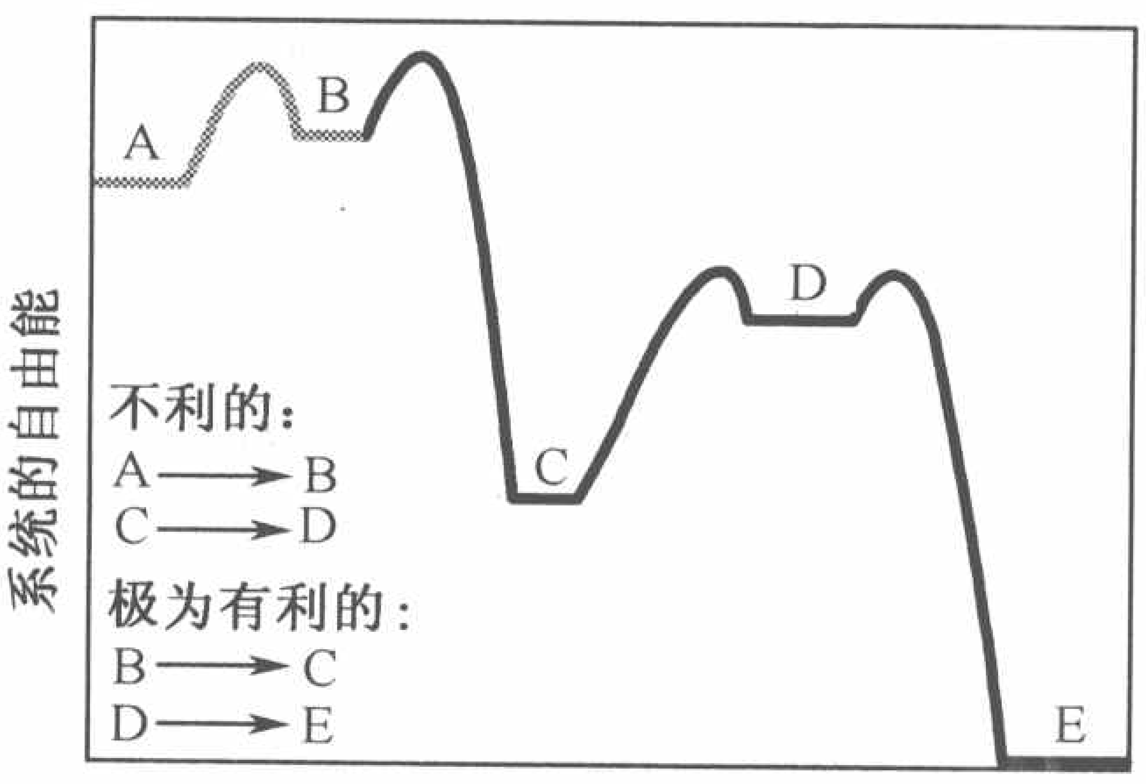


图 4-3　一个多步新陈代谢路径的自由能变化，A→B→C→D→E。其中两步（A→B 和 C→D）不利于 A→E 的反应方向，因为它们具有微正的 ΔG。然而由于步骤 B→C 和 D→E 所提供的极大的负 ΔG，它们已经没有意义了。所以，整个反应有利于 A→E 方向。

生物合成反应路径中并非每一步都需要高能键的断裂，经常仅仅是其中的一或两步需要。有时这是因为即使缺少外加高能键，ΔG 还是有利于生物合成反应方向。另一种情况是 ΔG 等于零或者甚至可能略微显正值。然而，与高能键水解作用耦联时的 ΔG 相比，这种微小的正 ΔG 是微不足道的。而且正如我们在图 4-3 看到的，只有一个反应路径的所有自由能变化的总和才是有意义的。如果某一特定生物合成反应步骤的后继步骤的 K_{eq} 是 100∶1，有利于生物合成反应的正方向，那么就不必再考虑它的

K_{eq}略微（80∶20）有利于降解反应。

同样地，降解反应路径中也并非每一步都产生高能键。例如，在葡萄糖漫长的酵解（Embden-Meyerhof）过程中只有两步产生 ATP。而且，许多降解反应路径中的一步或多步需要高能键的断裂。葡萄糖的酵解又是一个这样的例子，每产生 4 分子 ATP，它必须消耗 2 分子 ATP。当然，所有的产能降解反应过程都会产生比消耗的能量多的能量。

肽键的自发水解

由两个氨基酸生成一个二肽和一个水分子需要 1～4kcal/mol 的 ΔG，这依赖于参加反应的氨基酸类型。这种正值的 ΔG 告诉我们，多肽链不可能由自由氨基酸生成。更进一步，我们必须考虑这样一个事实，即细胞内水分子的浓度比其他任何分子高得多（一般情况下要高出 100 倍），所有水分子参与的反应的平衡都强烈地倾向于消耗水分子的反应方向。这可以由平衡常数的定义中简单看出。例如，生成二肽的反应：

$$\text{氨基酸(A)} + \text{氨基酸(B)} \rightarrow \text{二肽(A-B)} + H_2O \qquad [\text{方程 4-4}]$$

具有以下平衡常数：

$$K_{eq} = (\text{conc}^{A-B} \times \text{conc}^{H_2O})/(\text{conc}^{A} \times \text{conc}^{B}) \qquad [\text{方程 4-5}]$$

在这里浓度都是以 mol/L 做单位。所以，对于一个给定的 K_{eq}（方程 $\Delta G = -RT\ln K_{eq}$ 将其与 ΔG 联系在一起），大得多的水分子浓度意味着相对小的二肽浓度，而且相对浓度非常重要。实际上，简单的计算表明即使非水解反应的 ΔG 为 -3kcal/mol，水解反应仍可能自发进行。

所以，在理论上蛋白质是不稳定的，如果给以足够长的时间，蛋白质将会自发降解为自由氨基酸。并且，当缺乏特定的酶时，这种自发反应的速度太低以至于不能对细胞的新陈代谢产生轻微的影响。也就是说，蛋白质一旦形成，除非被特定的酶催化降解，否则都将保持稳定状态。

负值 ΔG 与正值 ΔG 的耦联

在能够相互结合形成蛋白质之前，氨基酸必须接受自由能。人们发现 ATP 作为能量供体的基本作用之后，就明白了这一过程是如何实现的。ATP 含有 3 个与腺嘌呤分子（腺嘌呤—O—Ⓟ～Ⓟ～Ⓟ）相连的磷酸基团，当一个或两个末端～P 基团被水解断裂时，自由能会显著下降。

腺嘌呤—O—Ⓟ～Ⓟ～Ⓟ＋H_2O→腺嘌呤—O—Ⓟ～Ⓟ＋Ⓟ　［方程 4-6］
（$\Delta G = -7$kcal/mol）

腺嘌呤—O—Ⓟ～Ⓟ～Ⓟ＋H_2O→腺嘌呤—O—Ⓟ＋Ⓟ～Ⓟ　［方程 4-7］
（$\Delta G = -8$kcal/mol）

腺嘌呤—O—Ⓟ～Ⓟ＋H_2O→腺嘌呤—O—Ⓟ＋Ⓟ　［方程 4-8］
（$\Delta G = -6$kcal/mol）

所有这些断裂反应都具有负值 ΔG，其绝对值（数字大小不考虑符号）要大于单体分子形成高分子时的正值 ΔG。自身具有正值 ΔG 的生物合成反应之所以能够进行，是因为它们与高能键的断裂反应相耦联，而后者伴随着具有绝对值更大的负值 ΔG。所

以，在蛋白质的合成过程中，每个肽键的形成（$\Delta G=+0.5$kcal/mol）都与 ATP 断裂为 AMP 和磷酸盐的反应相耦联，后者具有 -8kcal/mol 的 ΔG（方程 4-7）。这导致净 -7.5kcal/mol的 ΔG 已经足以保证平衡有利于蛋白质的合成而非断裂。

基团转移反应中前体的活化

当 ATP 被水解为 ADP 和磷酸盐的时候，大部分自由能转变为热。由于热能不能被利用形成共价键，所以两个分别具有正值 ΔG 和负值 ΔG 的完全分离的反应不能形成耦联反应。相反，耦联反应是由两个或多个连续反应构成。这些反应通常是基团转移（group-transfer）反应：在此类反应中，分子交换其功能基团（不包括氧化还原反应）。催化这类反应的酶被称作转移酶（transferase）。请看下面的反应：

$$(A-X)+(B-Y)\rightarrow(A-B)+(X-Y) \qquad [方程\ 4\text{-}9]$$

在这个例子中，X 基团与 B 相交换。转移基团反应被严格定义为不包括水分子参与反应。当有水分子参与时，

$$(A-B)+(H-OH)\rightarrow(A-OH)+(B-H) \qquad [方程\ 4\text{-}10]$$

这时反应就被称为水解反应，而参与的酶被称为水解酶。

在此，令我们感兴趣的基团转移反应是那些与高能键连接的基团。当这类高能基团被转移至适当的受体分子时，它们将与受体分子以高能键连接。因此，基团转移反应允许高能键从一个分子向另一个分子的转移。例如，方程 4-11 和 4-12 展示了 ATP 的能量是如何转移以形成 GTP 的，后者是 RNA 合成的前体之一：

腺嘌呤—Ⓟ～Ⓟ～Ⓟ＋鸟嘌呤—Ⓟ → 腺嘌呤—Ⓟ～Ⓟ＋鸟嘌呤—Ⓟ～Ⓟ

［方程 4-11］

腺嘌呤—Ⓟ～Ⓟ～Ⓟ＋鸟嘌呤—Ⓟ～Ⓟ → 腺嘌呤—Ⓟ～Ⓟ＋鸟嘌呤—Ⓟ～Ⓟ～Ⓟ

［方程 4-12］

GTP 的高能Ⓟ～Ⓟ 基团使其可以自发与其他分子连接。GTP 是所谓活化分子（activated molecule）的一个例子；类似地，转移高能基团的过程被称为基团活化（group activation）。

基团转移中 ATP 的多变性

在对作为能量供体的分子的能量所进行的调控中，ATP 的合成起着关键作用。不论是在氧化反应中还是光合磷酸化反应中，能量被用来从 ADP 和磷酸盐合成 ATP：

腺嘌呤—Ⓟ～Ⓟ＋Ⓟ＋能量 → 腺嘌呤—Ⓟ～Ⓟ～Ⓟ　［方程 4-13］

由于 ATP 是生物体中高能基团的首先接受者，因此它必然是一系列反应的起始点。在这些反应中，高能基团被转移到低能量分子以使其获得自发反应的潜能。ATP 的中心作用源于这样一个事实，即它含有两个高能键，每一个的断裂释放不同的基团。这可以由图 4-4 看出，图中展示了由 ATP 产生的 3 种重要的基团：Ⓟ～Ⓟ，焦磷酸基团；～AMP，单磷酸腺苷基团；～Ⓟ，磷酸基团。值得注意的是，只有当这些高能基团转移到适当的受

体分子上时，它们才显示出高能特性。例如，～Ⓟ 基团转移到 COO^- 基团会产生高能 COO～Ⓟ 磷酰基团，但是同样的基团转移到糖羟基（—C—OH）(正如 6-磷酸葡萄糖形成过程中发生的)，却只能产生低能量化学键（水解中产生的 ΔG 降低小于 5kcal/mol）。

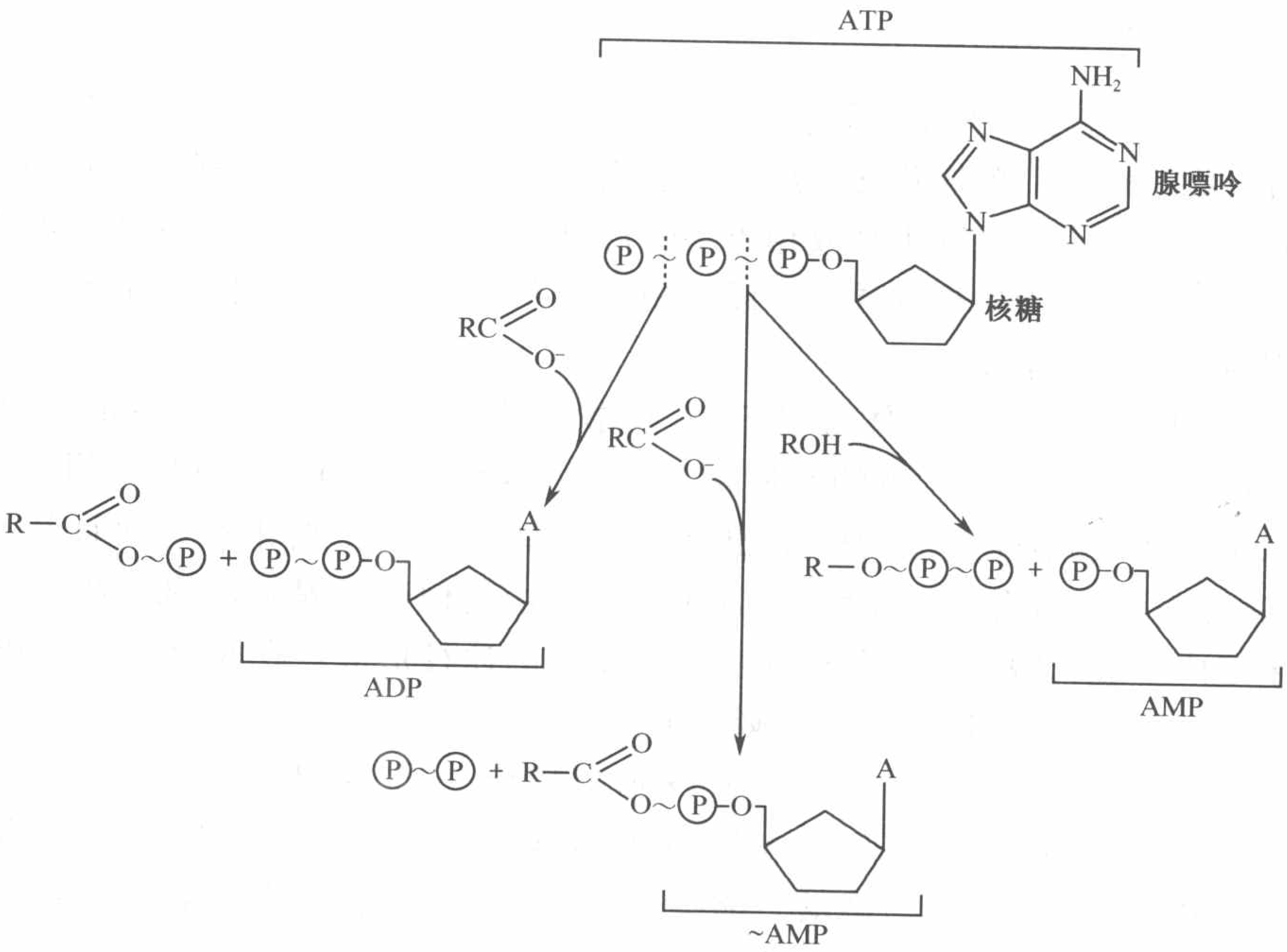
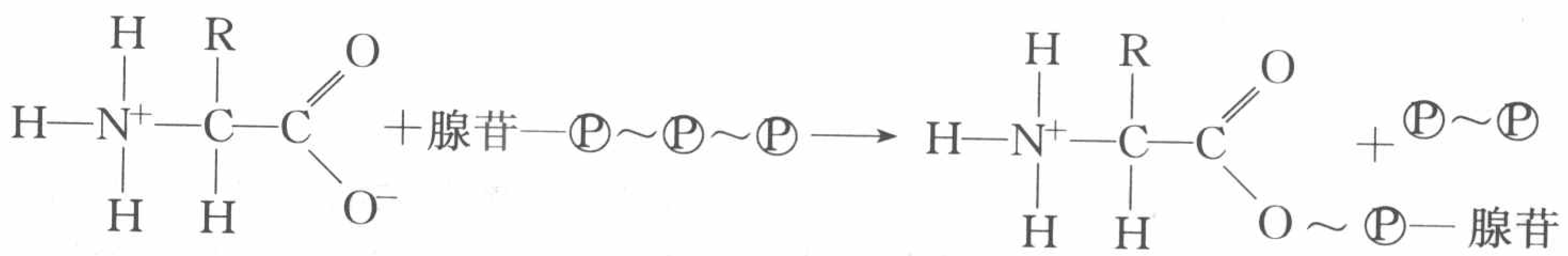


图 4-4 ATP 参与的重要基团转移反应。

与 AMP 连接使氨基酸活化

正如方程 4-14 所示，当 ATP 中的 AMP 基团转移到氨基酸的 COO^- 基团上时，氨基酸就被活化了：



［方程 4-14］

(方程中 R 表示氨基酸的特定侧链。）催化这一类型反应的酶称作氨基酸酰基合成酶(aminoacyl synthetase)。从热力学意义上讲，氨基酸（AA）一旦被活化，就能够用来合成蛋白质。虽然如此，AA～AMP 复合物并不是蛋白质的直接前体。相反，由于某种原因（我们将在第 14 章讨论）必须发生第二次转移基团反应，将羰基活化的氨基酸转移到一个 tRNA 分子的末端：

AA～AMP＋tRNA→AA～tRNA＋AMP ［方程 4-15］

通过 AA～tRNA 分子的浓缩作用，一个肽键在不断延伸的多肽链末端形成：

AA～tRNA＋延伸的多肽链(n 个氨基酸)→tRNA＋延伸的多肽链(n＋1个氨基酸)

［方程 4-16］

所以，正如和所有其他耦联反应一样，这一“耦联反应”的最终一步必须是移去活化基团和高能键转变为具有低水解自由能的化学键。这是驱动蛋白质合成反应方向的负值 ΔG 的来源。

核酸前体被Ⓟ～Ⓟ活化

不论是 DNA 还是 RNA，所有类型的核酸都是由单核苷酸单体（也被称为磷酸化核苷）所组成。然而，单核苷酸在热力学上比氨基酸更难聚合。这是因为将前者连接起来的磷酸二酯键在水解过程中释放相当多的自由能（－6kcal/mol）。这意味着核酸将以极低的速度自发水解为单核苷酸。所以，相对于在蛋白质的生物合成，在核酸的生物合成中使用活化前体更重要。

DNA 和 RNA 的直接前体都是 5′-三磷酸核苷。对于 DNA，这些前体是 dATP、dGTP、dCTP 和 dTTP（d 指的是脱氧）；对于 RNA，前体是 ATP、GTP、CTP 和 UTP。因此，ATP 不仅仅作为转基团反应中高能基团的主要来源，而且其自身就是 RNA 的直接前体。所有其他 3 种 RNA 前体都是由类似于方程 4-11 和方程 4-12 所描述的转基团反应生成。脱氧三磷酸核苷由几乎相同的途径形成：在脱氧单核苷酸合成后，它们通过从 ATP 转移基团转变为三磷酸形式。

脱氧核苷—Ⓟ＋ATP → 脱氧核苷—Ⓟ～Ⓟ＋ADP　［方程 4-17］

脱氧核苷—Ⓟ～Ⓟ＋ATP → 脱氧核苷—Ⓟ～Ⓟ～Ⓟ＋ADP［方程 4-18］

然后这些三磷酸核苷就能通过磷酸二酯键连接在一起形成多聚核苷酸，在这一转移基团反应中，一个焦磷酸键被打断，释放一个焦磷酸基团：

脱氧核苷—Ⓟ～Ⓟ～Ⓟ＋延伸的多聚核苷酸链(n 个核苷酸)

→ Ⓟ～Ⓟ＋延伸的多聚核苷酸链(n＋1 个核苷酸)［方程 4-19］

与形成肽键的反应不同，这一反应不具有负值 ΔG。实际上，ΔG 略微偏正（大约 0.5kcal/mol）。这种情况立即引出一个问题，因为多聚核苷酸显然在不断形成，它必需的自由能来源是什么？

核酸合成过程中释放的Ⓟ～Ⓟ能量值

多聚核苷酸形成过程中所需的自由能来自高能焦磷酸基团的分裂，后者与高能磷酸二酯键同时形成。所有的细胞都具有一种有效的酶——焦磷酸酶，这种酶会将刚刚形成的焦磷酸分子打断：

Ⓟ～Ⓟ → 2Ⓟ　　(ΔG ＝－7kcal/mol)　　［方程 4-20］

大的负值 ΔG 意味着反应是不可逆的。也就是说，一旦Ⓟ～Ⓟ断裂就不可能再次形成。

与焦磷酸基团的分裂［方程 4-20］相耦联的单磷酸核苷基团的聚合［方程 4-17］具有一个平衡常数，它由两个反应总的 ΔG 决定：(0.5kcal/mol)＋(－7kcal/mol)。最终的数值（ΔG＝－6.5kcal/mol）告诉我们，核酸几乎从不断裂生成其三磷酸核苷前体。

在这里我们看到了一个有力的实例，正是伴随一系列反应的总的自由能变化决定了其中某一反应是否能发生。具有微小正值 ΔG 的反应自身是不会自发发生的，但它们经常是重要的新陈代谢路径的一部分，其后继反应具有大的负值 ΔG。不论何时我们都应

该记住，一个单独反应（或甚至是一个单独路径）从不孤立存在；相反，通过代谢产物的加入和移去，反应的平衡在不断改变。

Ⓟ～Ⓟ 分裂是大多数生物合成反应的特征

核酸的合成并非唯一一个由Ⓟ～Ⓟ 的释放和分裂决定其方向的反应。实际上，所有的生物合成反应具有一步或多步释放焦磷酸基团的反应。以氨基酸与 AMP 结合活化为例，从其自身来说，高能键从 ATP 转移到 AA～AMP 复合物的过程具有微小的正值 ΔG。然而，正是 ATP 末端焦磷酸基团的释放和分裂提供了驱动反应所需的负值 ΔG。当我们考虑这样一个问题，即为什么细胞试图由三磷酸核苷合成核酸而不是单磷酸核苷（图 4-5），焦磷酸基团分裂的巨大作用就很清楚了。磷酸，而不是焦磷酸，被作为形成磷酸二酯键的骨架。然而，在有大量磷酸盐存在的情况下磷酸二酯键并不稳定，这是因为它们形成时没有释放大量自由能。所以，生物合成反应可被轻易逆转；如果磷酸盐大

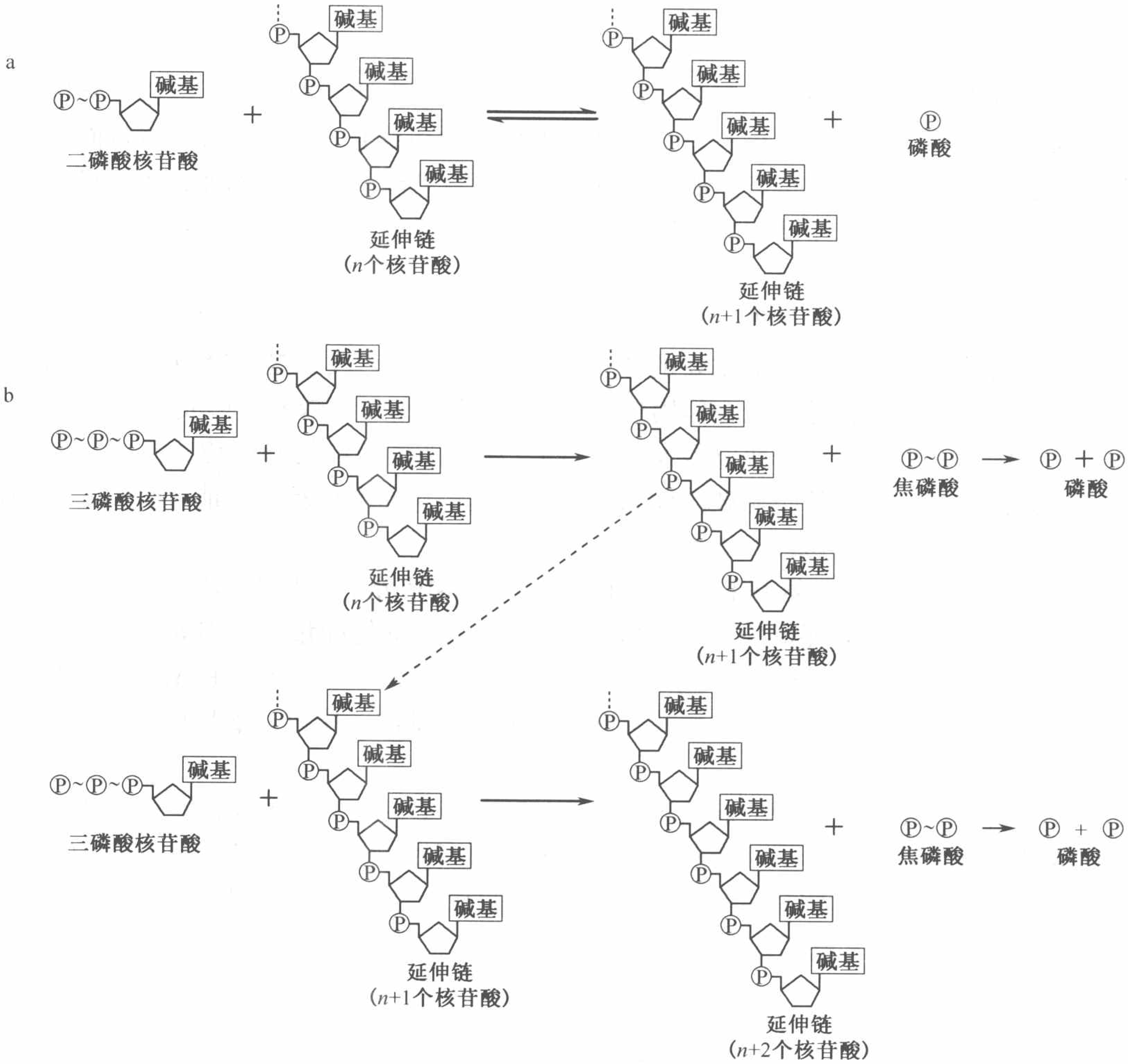


图 4-5　核酸生物合成的两种可能途径。 a. 运用二磷酸核苷合成核酸；b. 运用三磷酸核苷合成核酸。

量累积，根据质量反应定律，反应将开始向核酸分解的方向移动。而且对细胞来说，磷酸盐一形成就将其移除是不可行的（从而阻止逆向反应），因为所有细胞都需要一定浓度的内部磷酸盐维持生长。相反，产生焦磷酸然后迅速将其打断为两分子磷酸盐的一系列反应，消除了磷酸盐的产生和核酸的生物合成间的联系，从而阻止了逆转生物合成反应的可能性（图 4-5）。这样的结果是，细胞内很难积累足够的磷酸盐以驱动逆反应方向（或断裂方向）的两个反应。由此我们可以清楚地知道，运用三磷酸核苷作为核酸的前体并非随机选择的结果。

类似的讨论将告诉我们，为什么在所有细胞内都是由 ATP 作为高能基团的关键供体，而不是 ADP。起初这种 ATP 偏爱在生化学家看来是绝对的。但是现在我们发现，许多运用 ADP 作为能量供体的反应同样存在，且保持较好的反应方向。

小结

粗略看来，许多分子的生物合成好像违背了热力学定律，即自发反应必须伴随有自由能的下降（ΔG 是负值）。例如，从氨基酸形成蛋白质就具有正的 ΔG。当我们认识到生物合成反应并非如起初想像的那样进行时，问题自然就解决了。例如，蛋白质并非由自由氨基酸直接形成，而是首先被酶催化转化为高能活化分子的前体，在特定酶的催化下聚合形成所需的生物合成产物。

所以，许多生物合成过程是“耦联”反应的结果，第一个反应为第二个反应提供自发进行所需的能量。细胞内最主要的能量来源是 ATP，它或者是在降解反应（如发酵或呼吸）过程中，或者是在光合作用过程中，由 ADP 和无机磷酸盐形成的。ATP 具有数个高能键，在水解时产生数值很大的负 ΔG。与高能键相连的基团称为高能基团。通过基团转移反应，高能基团能被转移到其他分子上，从而形成新的高能化合物。这些衍生的高能分子是许多生物合成步骤的直接前体。

结合 AMP 基团后，氨基酸被活化形成 AA～AMP 分子，其中 AMP 来自于 ATP。AA～AMP 分子中的高能键的能量类似于 ATP 中高能键的能量。虽然如此，基团转移反应还是由于高能 P～P 分子的存在才能进行完全。高能 P～P 分子在 AA～AMP 分子的形成过程中产生，随后被焦磷酸酶打断生成低能量基团。所以，逆反应 P～P＋AA～AMP→ATP＋AA 不能存在。

几乎所有的生物合成反应起始于 P～P 的释放。P～P 一旦形成，立即就被酶催化打断为两个磷酸盐分子，从而消除了逆向生物合成反应的可能性。P～P 分裂的巨大作用为细胞的主要能量供体是 ATP（而非 ADP）提供了一个解释。ADP 不能转移高能基团，并同时产生 P～P 基团作为副产物。

（时亮 译 吴东颖 刘国振 刘韧 校）

参考文献

Berg J., Tymoczko J.L., and Stryer L. 2006. *Biochemistry*, 6th ed. Freeman, New York.

Kornberg A. 1962. On the metabolic significance of phosphorolytic and pyrophosphorolytic reactions. In *Horizons in biochemistry* (ed. M. Kasha and B. Pullman), pp. 251–264. Academic Press, New York.

Krebs H.A. and Kornberg H.L. 1957. A survey of the energy transformation in living material. *Ergeb. Physiol. Biol. Chem. Exp. Pharmakol.* **49:** 212.

Nelson D.L. and Cox M.M. 2000. *Lehninger principles of biochemistry*, 3rd ed. Worth Publishing, New York.

Nicholls D.G. and Ferguson S.J. 2002. *Bioenergetics 3*. Academic Press, San Diego.

Purich D.L. (ed.) 2002. Methods in enzymology: Enzyme kinetics and mechanism: Detection and characterization of enzyme reaction intermediates. *Methods in enzymology*, vol. 354. Academic Press, San Diego.

Silverman R.B. 2002. *The organic chemistry of enzyme-catalyzed reactions*. Academic Press, San Diego.

Tinoco I. (ed.), Sauer K., Wang J.C. and Puglisi J.D. 2001. *Physical chemistry: Principles and applications in life sciences*, 4th ed. Prentice Hall College Division, Upper Saddle River, New Jersey.

Voet D., Voet J.G., and Pratt C. 2002. *Fundamentals of biochemistry*. John Wiley & Sons, New York.

第5章 弱键和强键决定大分子的结构

本章概要

•

分子间和分子内相互作用决定高度有序结构

•

氢键的排列方式决定蛋白质的特定构型

•

大多数蛋白质由两个或三个结构域装配而成

•

弱化学键决定蛋白质在 DNA 和 RNA 分子上的正确位置

•

变构：通过改变蛋白质的外形调节其功能

DNA、RNA 和蛋白质都是由简单小分子组成的高分子。正如在第 4 章谈到的，高分子的合成依赖于活化小分子的受控催化聚合。对于 DNA 和 RNA 来说，这些小分子是核苷酸（图 2-11）；对于蛋白质来说，这些小分子是 20 种氨基酸，后者由其活化中间产物供体 tRNA 提供。在这些分子链的组装过程中，每增加一个小分子都需要断裂多个高能键。对于所有这些分子，构成元件的顺序决定其遗传学和生物化学功能。

在决定这些高分子的结构和功能的各种因素中，弱化学键起着决定性的作用。RNA、DNA 和蛋白质的主要信息就是组成它们的共价连接小分子的顺序。然而，只有当这些高分子在其不同部分之间形成广泛的附加弱化学键之后，它们才形成了特定的外形并使其能够行使其功能。在第 3 章已经讨论过的氢键、离子键、疏水作用和范德华力，不但指导蛋白质形成严格的结合位点，而且指导 DNA 形成双螺旋结构。事实上除了少数特例以外，对生物高分子这类相互作用的破坏因素（如通过加热或去污剂），即使共价键不被破坏，都将会完全破坏其所有的生物活性。本章将会简略探讨生物大分子的结构以及保持其外形的作用力。DNA 和 RNA 在此仅做简要介绍，更多内容将在第 6 章讲述。我们将主要关注多种多样的蛋白质的结构。本章最后一节集中于蛋白质与核酸的相互作用，这是我们在本书中将要遇到的许多生物过程的中心环节，另外还有通过变构作用调控蛋白质的功能。

分子间和分子内相互作用决定高度有序结构

DNA 能够形成规则的螺旋

DNA 分子通常具有规则的螺旋构型，这是由于大多数 DNA 分子拥有两条具有互补结构的反向平行的多聚核苷酸单链（更多细节请参考第 6 章）。分子内和分子间的非共价键都起着稳定该结构的作用。两条单链由互补的嘌呤和嘧啶对之间的氢键连接在一起（图 5-1）。腺嘌呤总是与胸腺嘧啶形成氢键，而鸟嘌呤总是与胞嘧啶形成氢键。另外，实际上核糖和磷酸基团的所有表面原子都与水分子形成了氢键。

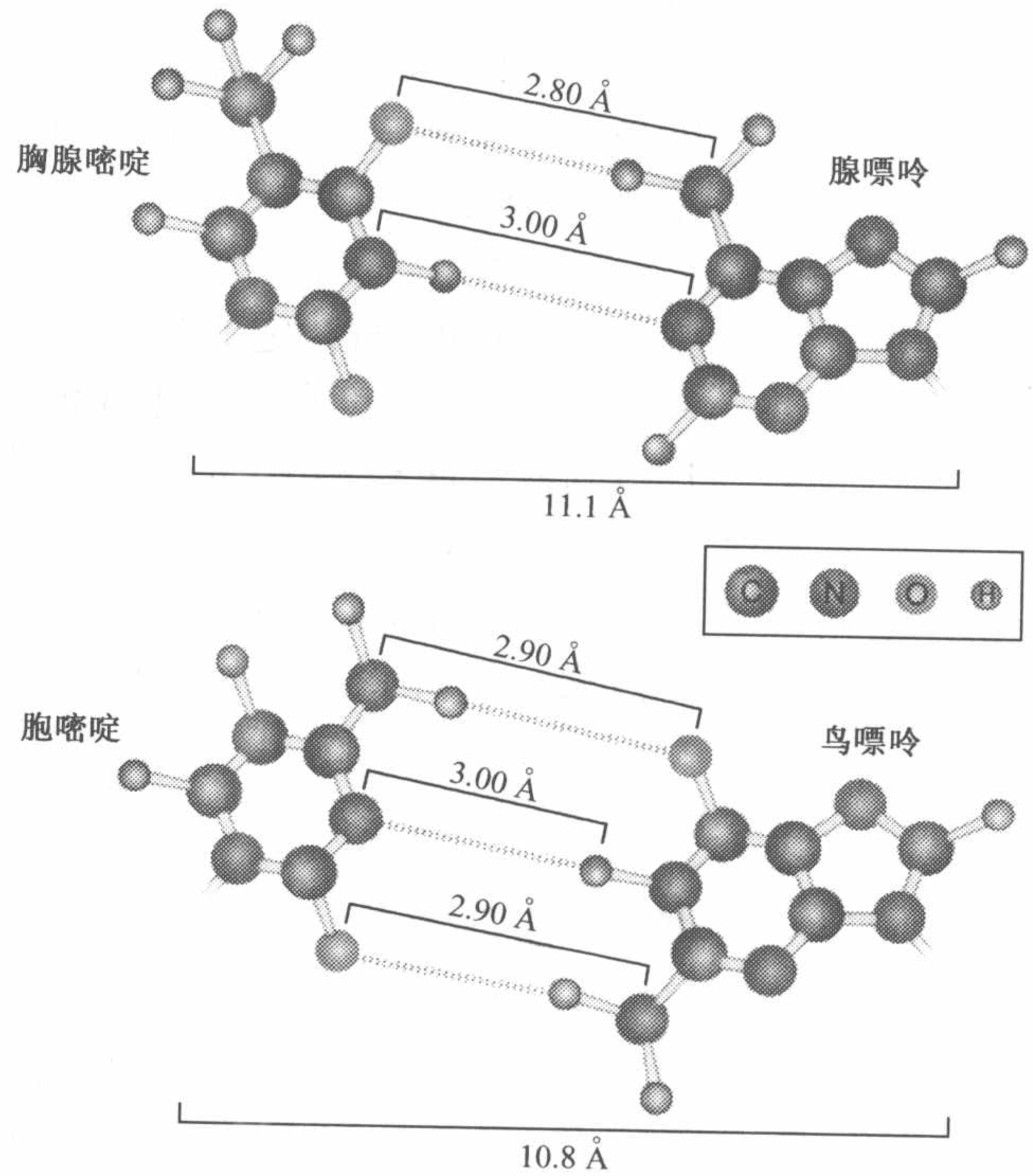


图 5-1 DNA 分子中以氢键相连的碱基对。图中展示了碱基对间氢键的位置和键长，每个碱基内部原子间的共价键也有标出，但没有区分单键和双键（见图 6-6）。

嘌呤嘧啶碱基对处于 DNA 分子的中心部位。这种排列允许它们的碱基平面依次层叠堆积，从而在碱基间形成共享（π-π）电子，阻止其与水分子结合。如果仅有一条多聚核苷酸链存在，碱基堆积排列将会被大大削弱。由于嘧啶分子小于嘌呤分子，单链 DNA 将会导致相邻碱基之间疏水表面的不利暴露，而双螺旋 DNA 中互补碱基对的存在使 DNA 能够形成规则结构，这是因为每个碱基对都是同样大小。

双螺旋 DNA 分子非常稳定有两个原因。首先，双螺旋结构的破坏将会导致疏水性的嘌呤和嘧啶与水分子形成更多的接触，这是非常不利的。其次，双螺旋 DNA 分子具

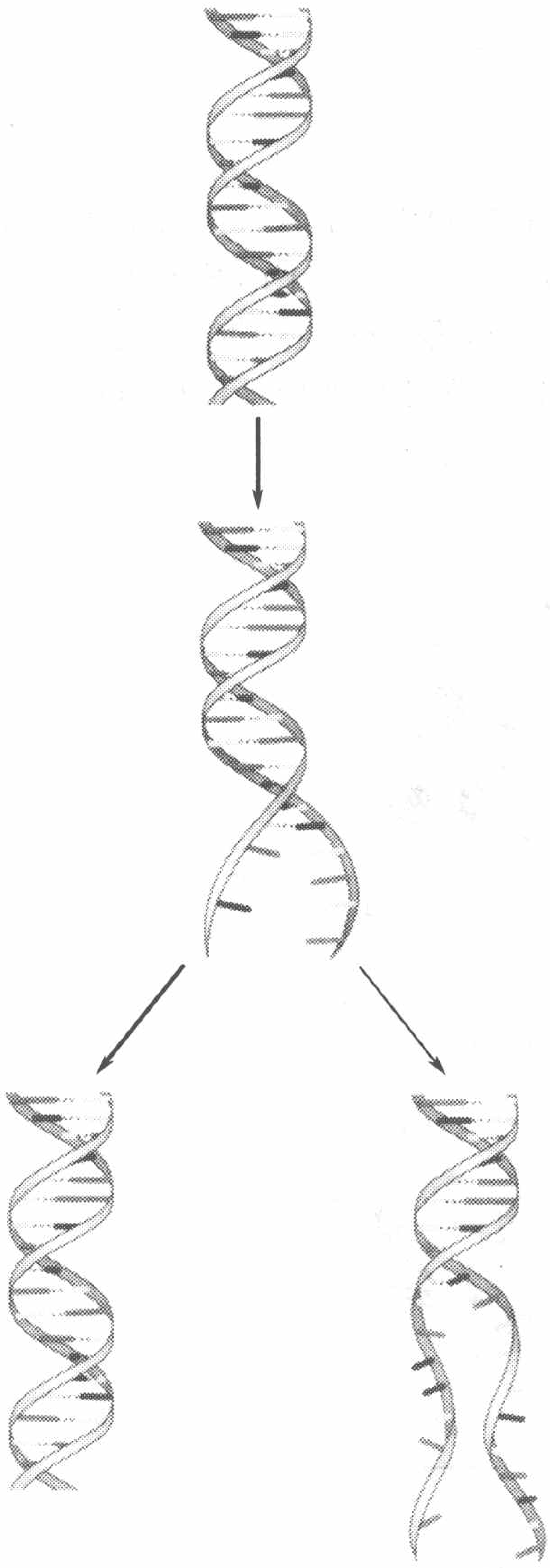


图 5-2　DNA 末端的碱基对被随机热运动打断。图中显示一旦末端的某些键被打断，可以重新形成（见下左），否则，其他键可被打断。

有非常多的弱化学键，在其排布状态下大多数弱化学键具有以下特性，即在其他许多化学键被同时破坏时，它们才有可能被破坏。例如，即使热运动不断在分子末端拆开嘌呤-嘧啶对，两条分子链并不因此而分开，这是因为分子中的其他氢键仍然紧密结合（图 5-2）。一旦某个氢键断裂，紧接着会进行该氢键的重建以维持原有的分子结构，而非更多键的断裂。当然，有时一个键断裂后紧随着第二个键，甚至更多键的断裂。然而，这类多个氢键断裂现象极为少见，所以由 10 个以上碱基对连接在一起的双螺旋 DNA 分子在室温下也很稳定。当双螺旋 DNA 分子确实分开而不重建时，这通常起始于分子的一端并向分子内部延伸。这是因为 DNA 分子末端碱基间的相互作用受邻近效应的影响最小，也就是说，它们只有一个邻近碱基对来帮助它们稳定相互作用。正如下面将要详细介绍的，多重弱化学键的原理也同样决定了蛋白质的稳定性。

当分子的温度高于生理温度时，次要化学键的有序排列就会越来越不稳定，随着温度的提高，一些弱化学键被同时打断的概率会逐渐上升。当一定数量的弱化学键被打断后，分子通常会失去其原有外形（变性过程）从而形成无活性的（或变性的）结构。因此，当温度上升时，DNA 分子需要更多的相互作用以保持双螺旋结构。

RNA 形成多种多样的结构

相对于 DNA 分子形成的高度规则的双螺旋结构，RNA 分子通常以单链分子形式存在。有些 RNA 分子（如 mRNA）作为遗传信息的临时携带者，不断与蛋白质之类的分子结合，它们不具有独立而稳定的三级结构。另外一些 RNA 分子折叠成特定的三级结构。对于 RNA，邻近区域间的分子内相互作用影响着二级结构元件的结构，这类相互作用主要存在于 RNA 的碱基之间，包括传统的 Watson-Crick 碱基配对、仅在 RNA 中被发现的异常碱基配对以及疏水性碱基堆积。RNA 和 DNA 的不同点在于组成其骨架的核糖携带有 2′-羟基。在 RNA 分子的折叠结构中，这些 2′-羟基经常参与相互作用以稳定结构。二价金属离子（如 Mg^{2+}、Mn^{2+} 和 Ca^{2+}）与 RNA 的结合经常是稳定的折叠构型形成的必要条件，这是因为金属离子能够屏蔽 RNA 骨架的负电荷，允许分子的不同区域更加接近。

某些重要 RNA 分子的高度可分解的结构是 RNA 三级结构精密而细致的折叠的最好说明，一种参与蛋白质合成的分子 tRNA（图 14-5）就是一个这样的例子。这些结构显示出碱基堆积在 RNA 构型中所起的主要作用。例如，tRNA 76 个碱基中的 72 个处于碱基堆积状态。正如在 DNA 双螺旋结构中一样，RNA 碱基的首尾相接堆积在能量上是有利的。因此，RNA 中由较少碱基对形成的短螺旋区域倾向于首尾相连，以形成较长的不连续螺旋区域，然后这些堆积的螺旋区域通过形成附加的三级结构相互靠近。

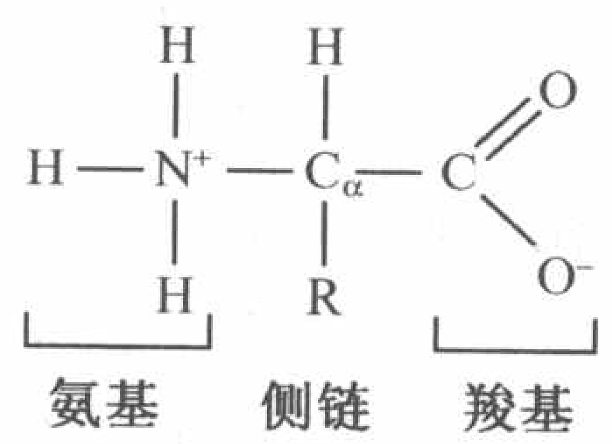


图 5-3　氨基酸的一般结构特性。

在此，我们仅仅简略的探讨了 DNA 和 RNA 的结构特性。我们将在第 6 章详细讨论在细胞中决定这些关键分子构型的相互作用。在本章的其余部分我们将会重点关注影响蛋白质结构的作用力。

蛋白质构成单元的化学特性

相对于组成 RNA 或 DNA 的 4 种核苷酸小分子，组成蛋白质的 20 种氨基酸小分子各不相同。氨基酸分子的一般结构特性是中心碳原子（C_α）连接着一个氢原子、一个基本的氨基以及一个羧酸基团（图 5-3），第四个连接者是被称作 R 基团（R group）的各种不同侧链。20 种氨基酸的 R 基团能够以大小、外形和化学成分加以分类(图 5-4)。

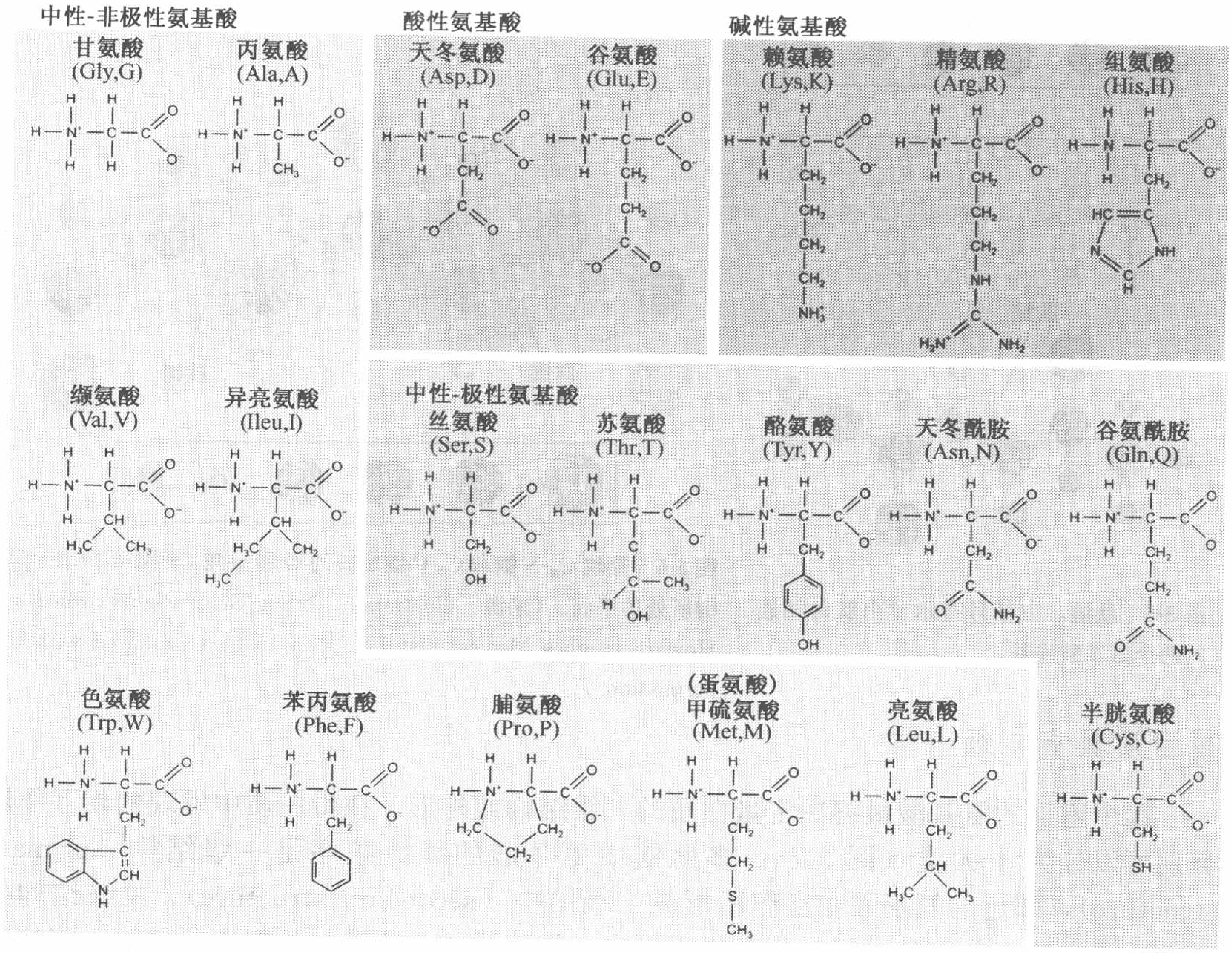


图 5-4　存在于蛋白质中的 20 种不同氨基酸。括号中为氨基酸的缩写形式（包括氨基酸的单字母表示方法）。

R 基团主要分为四类：中性-非极性、中性-极性、酸性和碱性。中性-非极性侧链由简单碳链或芳环组成，主要以疏水作用相互接触；中性-极性侧链包括羟基、巯基、酰胺基团和咪唑基团，主要形成氢键；带电（酸性和碱性）侧链包括基本的以及衍生的氨基和羧基，形成离子键和氢键。四种类型的侧链都参与范德华力相互作用，因为这种相互作用仅依赖于原子的邻近，而非其各自的化学组成（第 3 章）。

肽键

正如第 3 章图 3-2 所言，蛋白质中氨基酸间的主要共价键是肽键（peptide bond）（图 5-5）。当一个氨基酸的氨基与另一个氨基酸的羧基基团共价结合时就会形成肽键。这种化学键具有部分的双键特性。由于这种化学键包含不止一对共用电子，所以围绕其进行旋转是不允许的；只有当原子与一个单键连接时，围绕其进行的完全自由的旋转才有可能实现（例如，乙烷中的甲基基团，$H_3C—CH_3$，围绕 C—C 单键旋转）。相对于肽键而言，多肽骨架中的所有其他化学键都是单键，允许自由旋转。从热力学上来讲，这些化学键能够以无限种可能的构象存在；然而，在蛋白质的真实情况中，邻近肽段间的空间位阻限制了它们的旋转。处于同一平面的邻近肽键的定位可由两个键角描述：Φ 和 ψ（图 5-6）。在蛋白质中，多肽骨架中的功能基团间需要尽量多的形成次要化学键并同时尽量减小空间位阻，这一需要限定了以上两个键角。

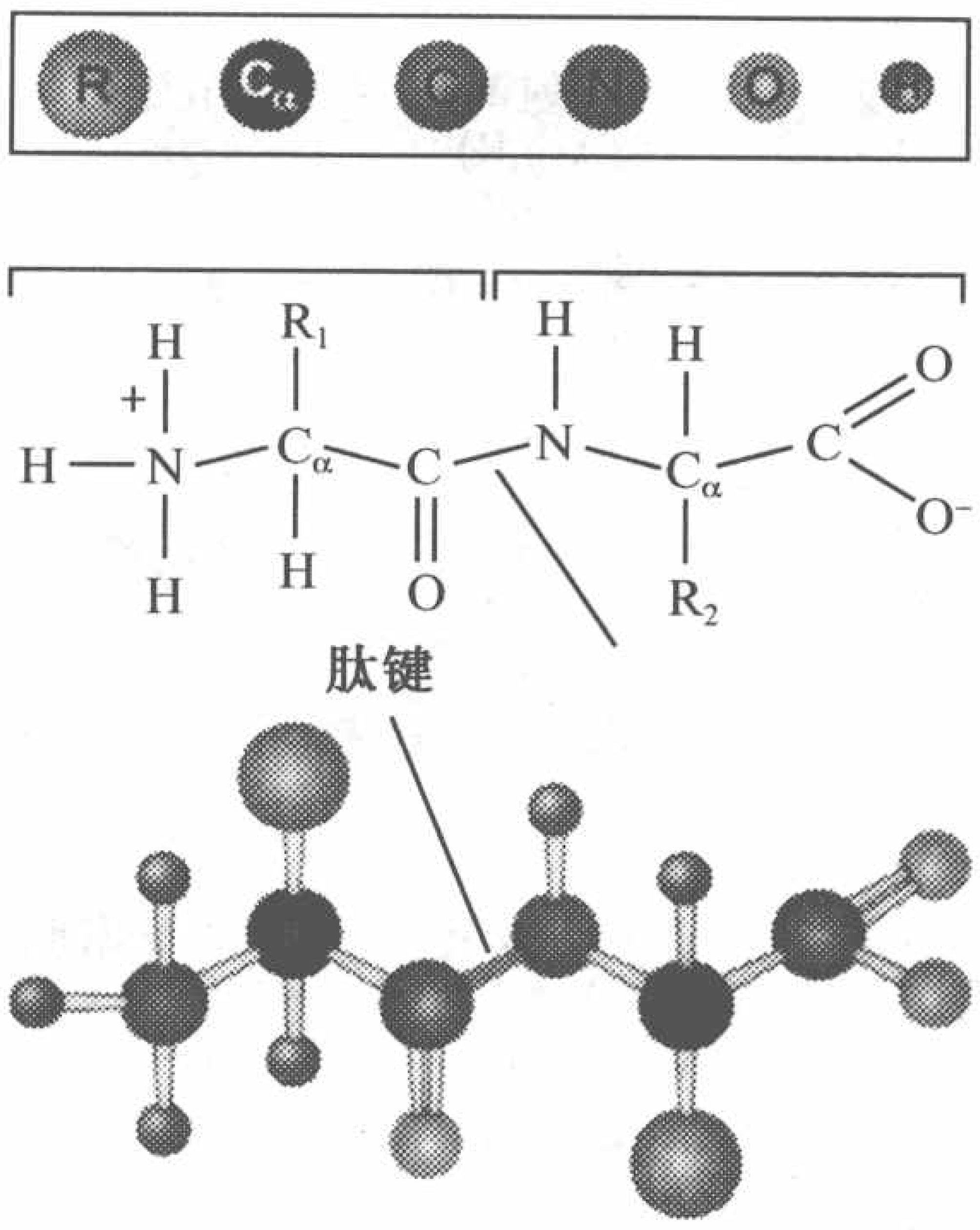


图 5-5 肽键。方括号指示出由肽键相连的两个氨基酸残基。

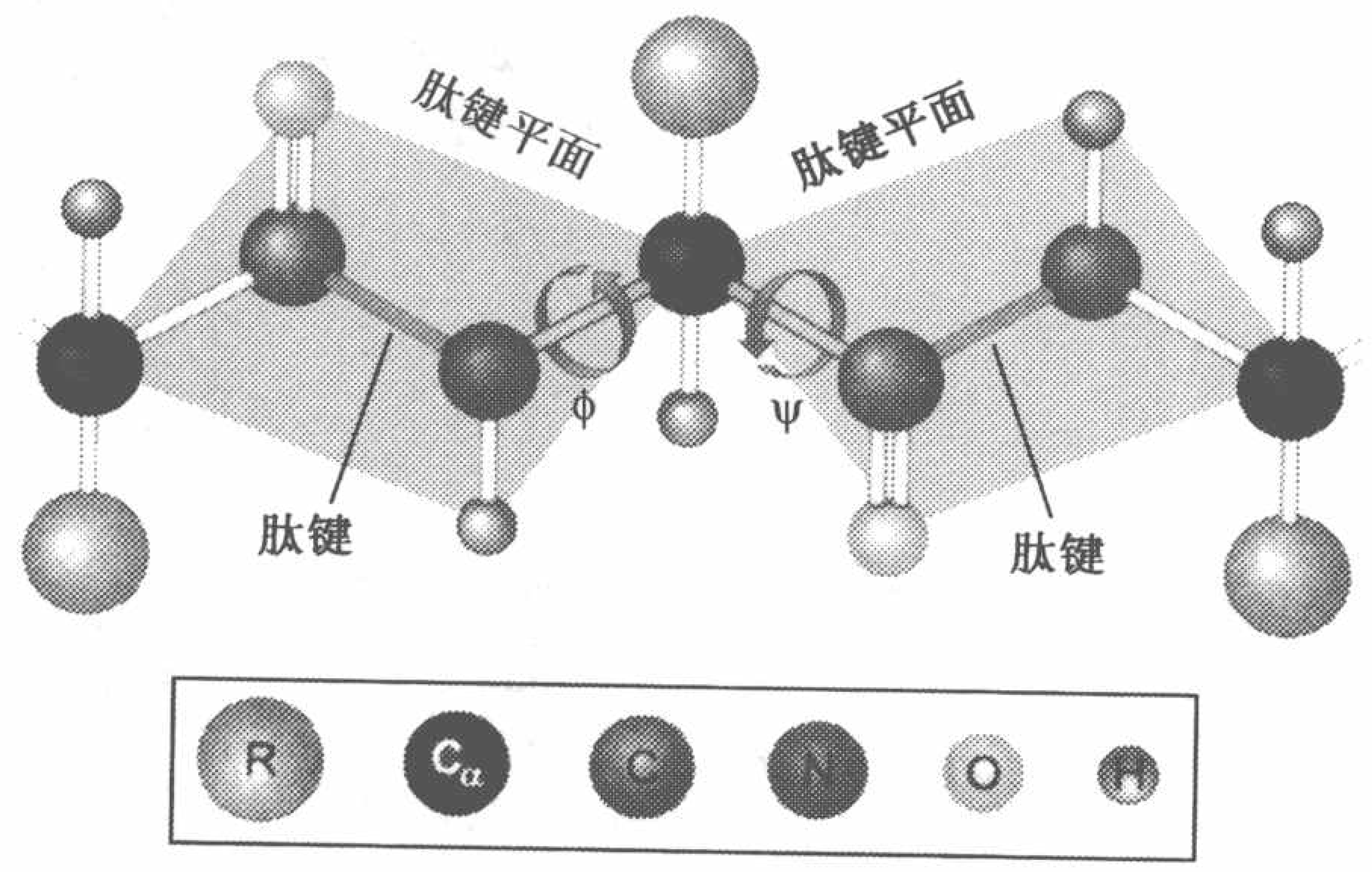


图 5-6 围绕 Cα-N 键和 Cα-C 键旋转的 Φ 和 ψ 角。阴影部分表示肽键所处的平面。（来源：illustration，Irving Geis. Rights owned by Howard Hughes Medical institute. Not to be reproduced without permission.）

蛋白质具有 4 级结构

逐个增加的氨基酸最终决定蛋白质的三维结构或外形。在蛋白质中发现的相互作用类别可以分为 4 大类（图 5-7）。多肽链中氨基酸的线性顺序是一级结构（primary structure），邻近的氨基酸相互作用形成二级结构（secondary structure）。二级结构的元件通常由这部分氨基酸间的相互作用形成，它们构成多肽链的骨架而非侧链。正如我们将要看到的，α 螺旋和 β 折叠就是二级结构元件。这些元件以特定的规则相互接近形

成给定多肽的三级结构（tertiary structure），它是单条多肽链的整体构型。许多蛋白质是由多条多肽链组成的，后者即蛋白质亚基（protein subunit），这些亚基相互连接的规则被看作蛋白质的四级结构（quarternary structure）。

多肽一级结构中所包含的信息通常已经足以决定其最终的三级结构，这一点已由一个经典实验所证明。在该实验中，一种单链多肽酶——核糖核酸酶被置于严格的条件下。在该条件下，氢键以及其他弱化学相互作用都受到影响，导致多肽的完全变性（或去折叠）。当变性的核糖核酸酶被重新置于允许形成弱化学键的条件下时，这种酶迅速恢复其正常三维结构和RNA切割活性。如何根据蛋白质结构设计实验的描述见框5-1。

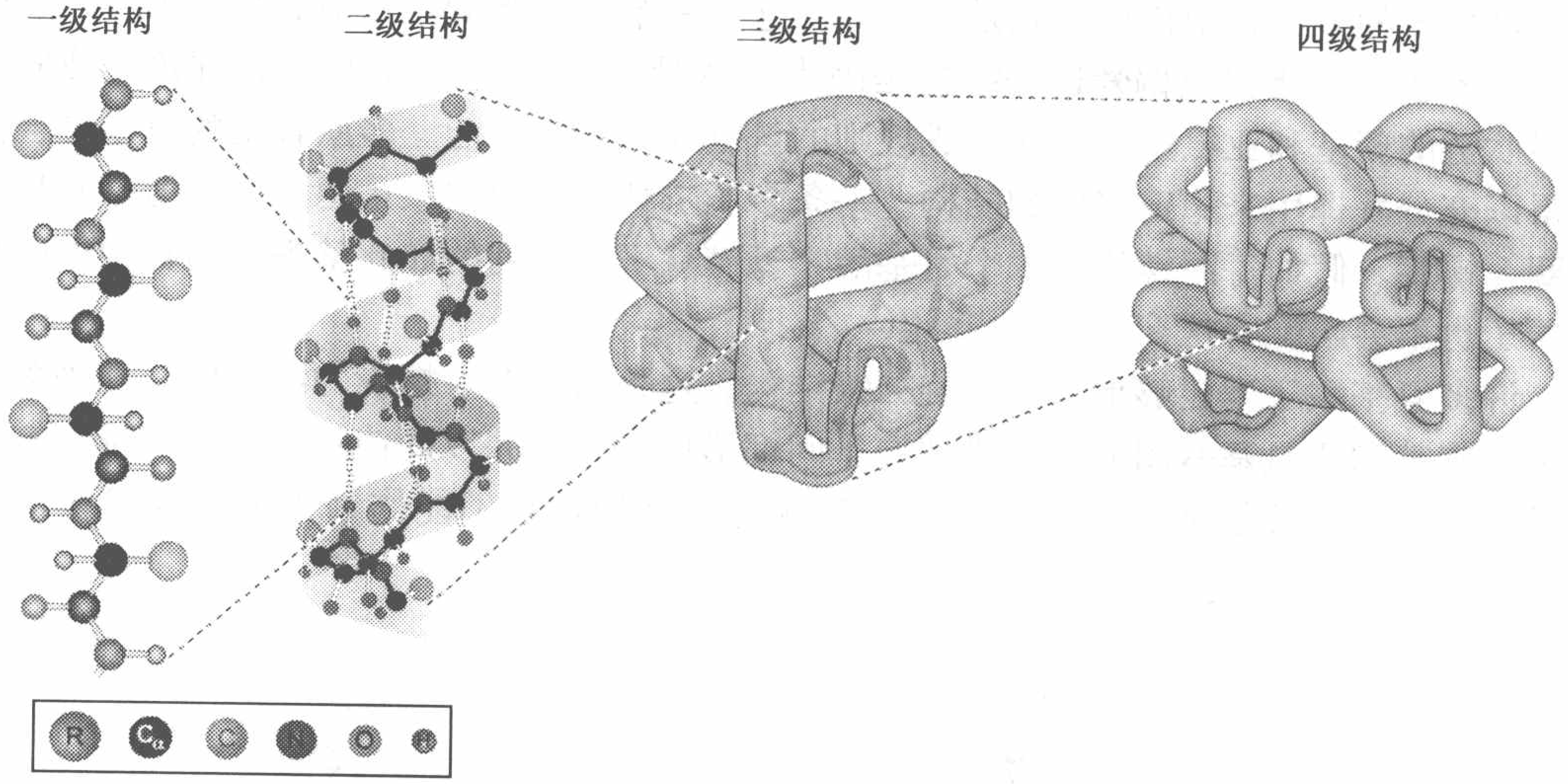


图 5-7 蛋白质结构的 4 个层次。（来源：Adapted from Branden C. and Tooze J. 1999 Introduction to protein structure，2nd edition，p. 3. Fig 1. 1.）

框 5-1 蛋白质结构的确定

确定蛋白质的三维结构主要有两种方法。首先发展起来的是X射线结晶学，这种方法依赖于纯蛋白质高度有序结晶的形成。正如在对DNA纤维所进行的最初的衍射研究中那样，用高能X射线照射蛋白质结晶会产生与蛋白质结构相关的衍射图案。最近，核磁共振技术已经被发展来判断更小的蛋白质分子的构型。这种技术利用中心原子（如 ^1H）的电磁性质检测邻近原子间是如何相互影响的。这些信息能够被用来判断多肽链中特定原子的相对位置，而这些距离就预示了蛋白质的整体结构（图5-7）。

从理论上来说，我们可以通过蛋白质的一级氨基酸序列来预测蛋白质的三维结构，因为这些信息终究已经足以使蛋白质形成唯一的构型。尽管基于氨基酸序列预测蛋白质结构的研究已经取得进展，对一段序列能态（energetic constraint）的完全判断仍然超出了最强大的计算手段的能力范围。然而，对特定二级结构元件（如下面将要介绍的一般的α螺旋结构）的预测已经越来越可靠。

用实验手段确定的蛋白质结构的数目的增加，为基于氨基酸序列预测蛋白质结构提

供了一个重要的信息。已有的原子结构能够帮助我们确定那些具有相似氨基酸序列的蛋白质的三维结构。通过将未知结构的蛋白质序列与那些已经被证实的序列进行比较，经常可以根据相似性做出结构预测。将这些信息与预测二级结构的计算机算法结合起来已经成为预测蛋白质是如何折叠的强有力的方法。远期展望告诉我们，这些手段将使由任一蛋白质的一级序列单独做出至少是近似的结构预测成为可能。

α 螺旋和 β 折叠是常见的二级结构形式

最稳定的多肽骨架排列方式是 α 螺旋。这是一个右手螺旋，沿着螺轴每 5.4Å 旋转一周（图 5-8）。蛋白质偏爱这一结构是因为多肽骨架具有适合的 Φ 和 ψ 键角，使同一条多肽链的羰基和亚氨基基团间形成规则的氢键模式成为可能。多肽骨架的氢键结合潜能得以充分发挥用来稳定多肽的结构。一条多肽链精确几何构象是由 3.6 个氨基酸形成一圈 α 螺旋。假如每圈螺旋有 4 个氨基酸，氢键就不会排列的如此整齐，而骨架中的单个原子也不会匹配得这么好。

许多氨基酸序列能够形成 α 螺旋二级结构。这是因为在 α 螺旋的结构中，几乎多肽链中的羰基和亚氨基基团的全部骨架原子间都形成连接，并以此使结构稳定。唯一缺乏这类原子的氨基酸是脯氨酸。由于它的环状化学结构，它不能作为供体参与氢键的形

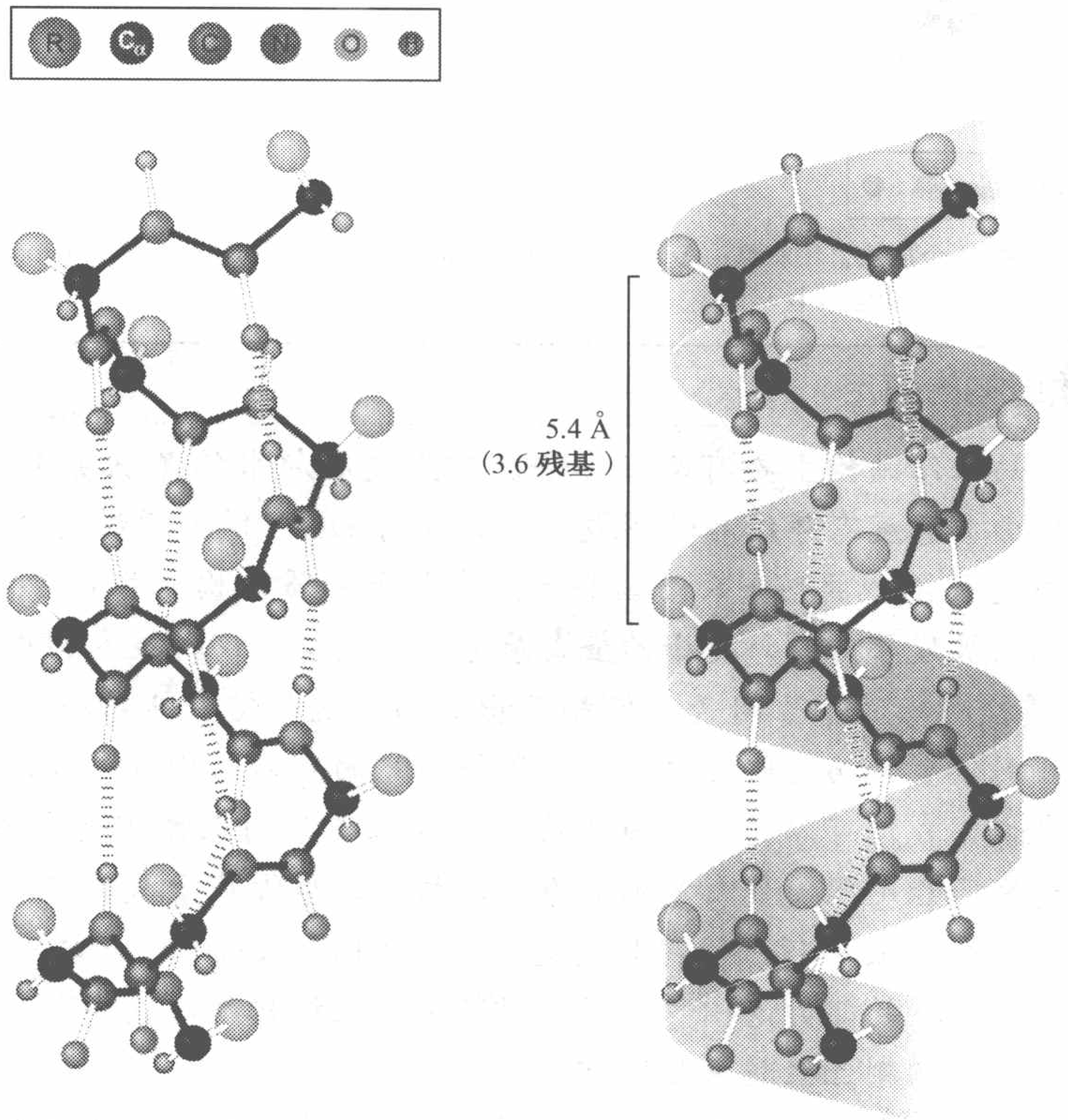


图 5-8 多肽链折叠形成 α 螺旋构象。（来源：Molecular structure adapted from Pauling L. 1960. The nature of the chemical bond and the structure of molecules and crystals: An introduction to modern structural chemistry, 3rd edition, p. 500. Copyright© 1960 Cornell University. Used by permission of the publisher.）

成，而后者正是稳定结构的关键。因此，脯氨酸是一个螺旋断裂残基（helix-breaking residue）。尽管甘氨酸、酪氨酸和丝氨酸的结构并不妨碍其形成α螺旋，但它们很少出现在α螺旋里。从α螺旋是由排他性的骨架间的连接维持这一事实得出的另一结果是，侧链将会由螺旋向四周伸展。这使得这些侧链能够处于一种理想状态，促进其与蛋白质的其他区域或另一个大分子（如DNA）的相互作用。

第二种常见的二级结构元件是β折叠（图5-9）。与α螺旋不同，β折叠是多肽骨架的一种高度延展形态。β折叠结构之所以稳定是因为在这种延展构型中多肽链区域整齐排列，使得一个β折叠单链的羰基能够与相邻单链的NH基团形成氢键。在典型情况下，β折叠的一个区域含有4～6条独立的多肽单链（每一条构成一个单独的β单链），每条有8～10个氨基酸的长度。在β折叠中，相邻氨基酸间具有180°的旋转角度，因此它们各自的侧链基团处于β折叠的异侧（图5-9b）。

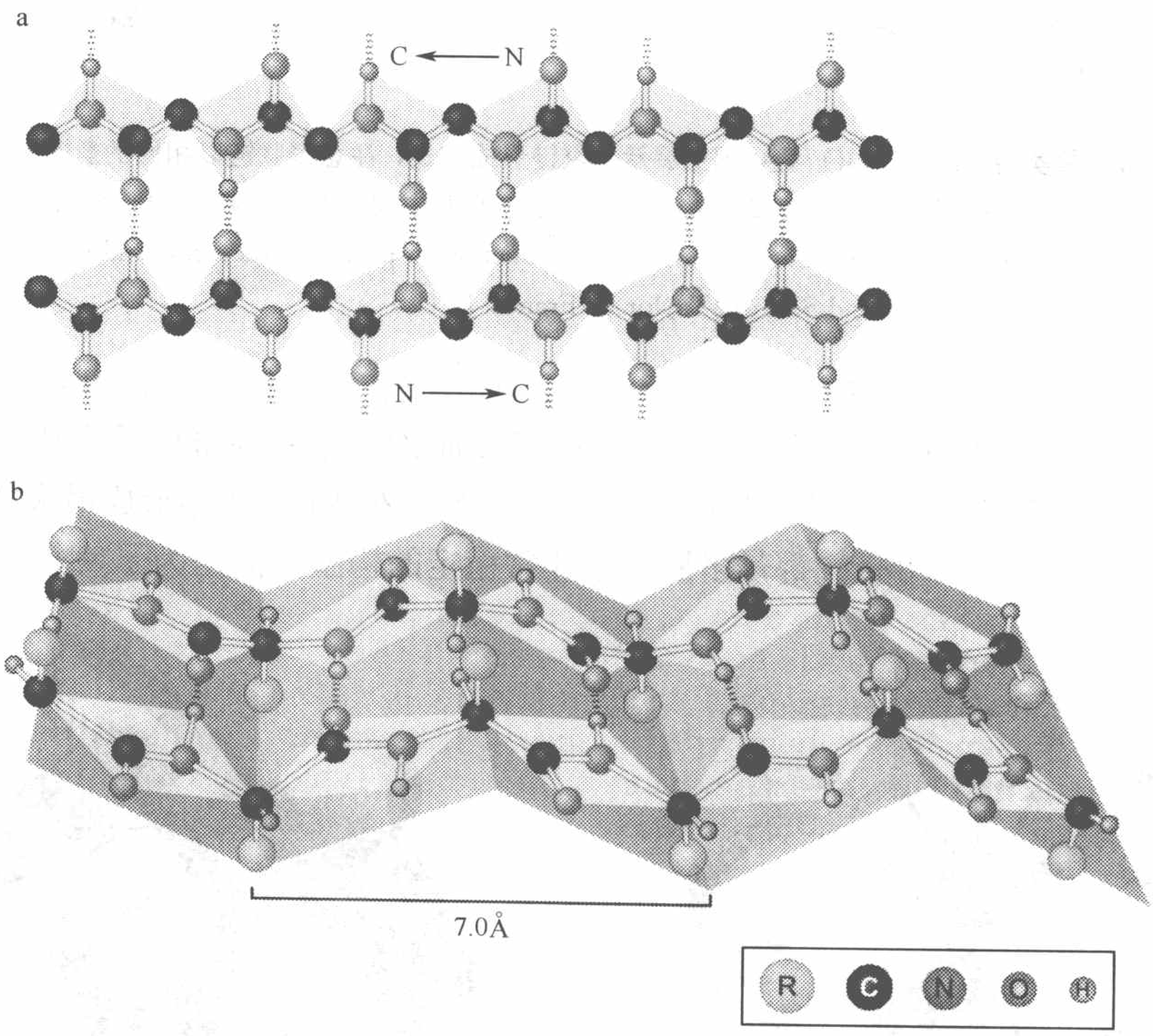


图5-9 由氢键结合形成的β折叠。 a. β折叠的俯视图。注意骨架中的氧原子和氮原子全都形成了氢键；b. β折叠的侧面图。该图展示了侧链基团的定位，它们交替处于β折叠平面的上方和下方。本图强调了β折叠的褶皱形状。这里显示的是反平行β折叠（见图5-10）。（来源：Illustration，Irving Geis. Rights owned by Howard Hughes Medical Institute. Not to be reproduced without permission.）

β折叠的两种形式中有一种占有绝对优势。这两种形式的区别在于多肽链的相对位置（图5-10）。在一种形式中，邻近多肽链以相同的氨端到羧端方向排列形成平行（parallel）β折叠；在另一种形式中，邻近多肽链以反方向排列构成反平行（antiparallel）β折叠。尽管不太常见，但确实存在同时具有平行和反平行部分的β折叠。不论是在平行还是反平行的β折叠中，所有多肽近似处于折叠的平面中。结构研究表明，在大

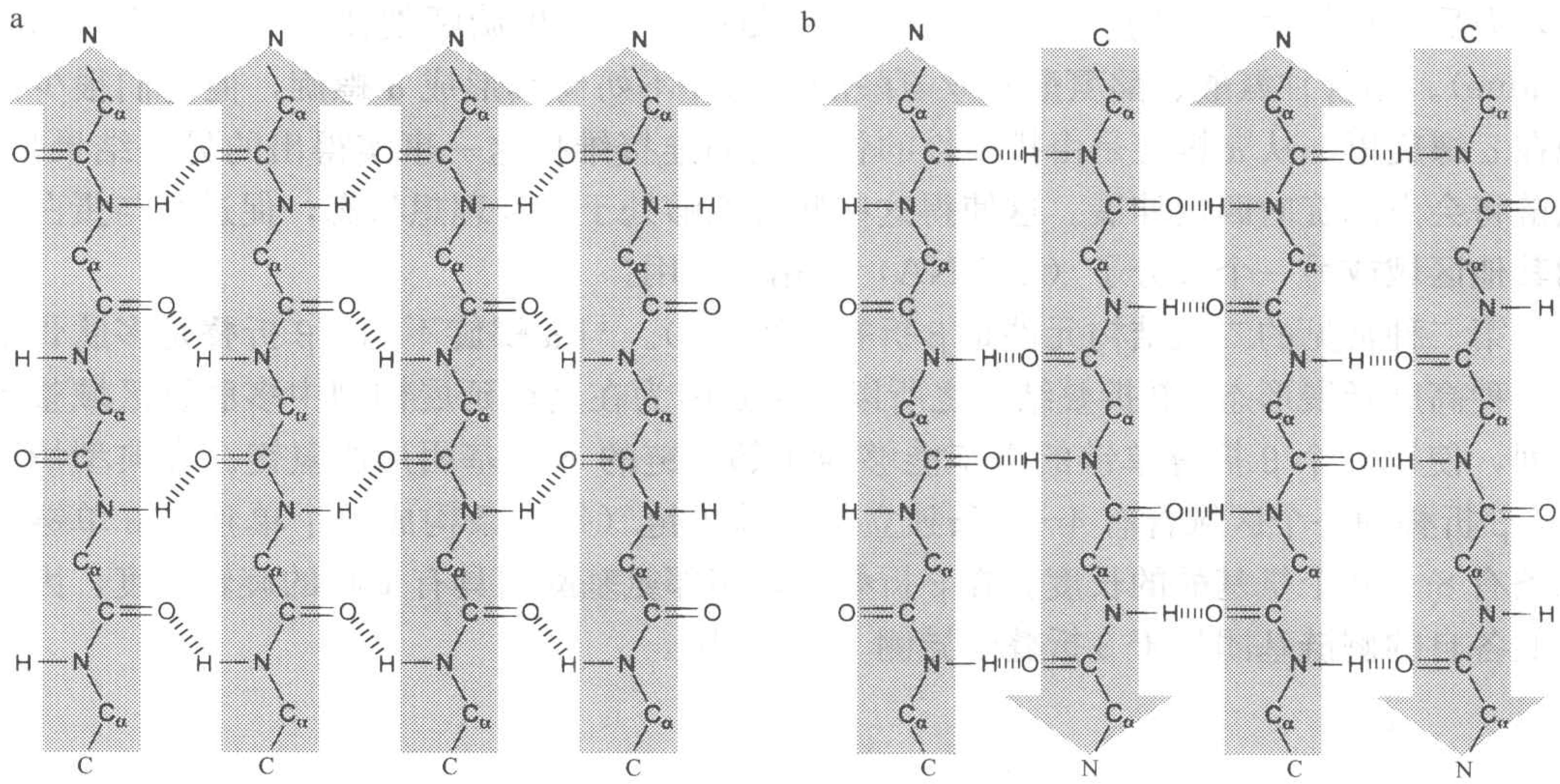


图 5-10 两种 β 折叠类型。 a. 平行 β 折叠：图上展示了氢键形成方式，注意多肽链以相同的氨基端到羧基端方向排列；b. 反平行 β 折叠：图上展示了氢键形成方式，注意 β 折叠中的基本 NH 和 O 原子都以氢键相互连接。（来源：Adapted from Branden C. and Tooze J. 1999. Introduction to protein structure，2nd edition，p. 19，Fig 2. 6a and p. 18，Fig 2. 5b. © Garland Science/Taylor & Francis LLC.）

多数情况下，β 折叠中的单独单链趋向于沿着长轴形成右手缠绕（图 5-11）。所以，β 折叠区域不是形成蛋白质的平面折叠，而是趋向于卷曲构成紧密的蛋白质模块。

为了使蛋白质能够正确折叠，蛋白质骨架和侧链都必须采取能够形成最大的有利的相互作用的构型。α 螺旋和 β 折叠都是多肽骨架非常稳定的构型。但是为了使每一个侧链都

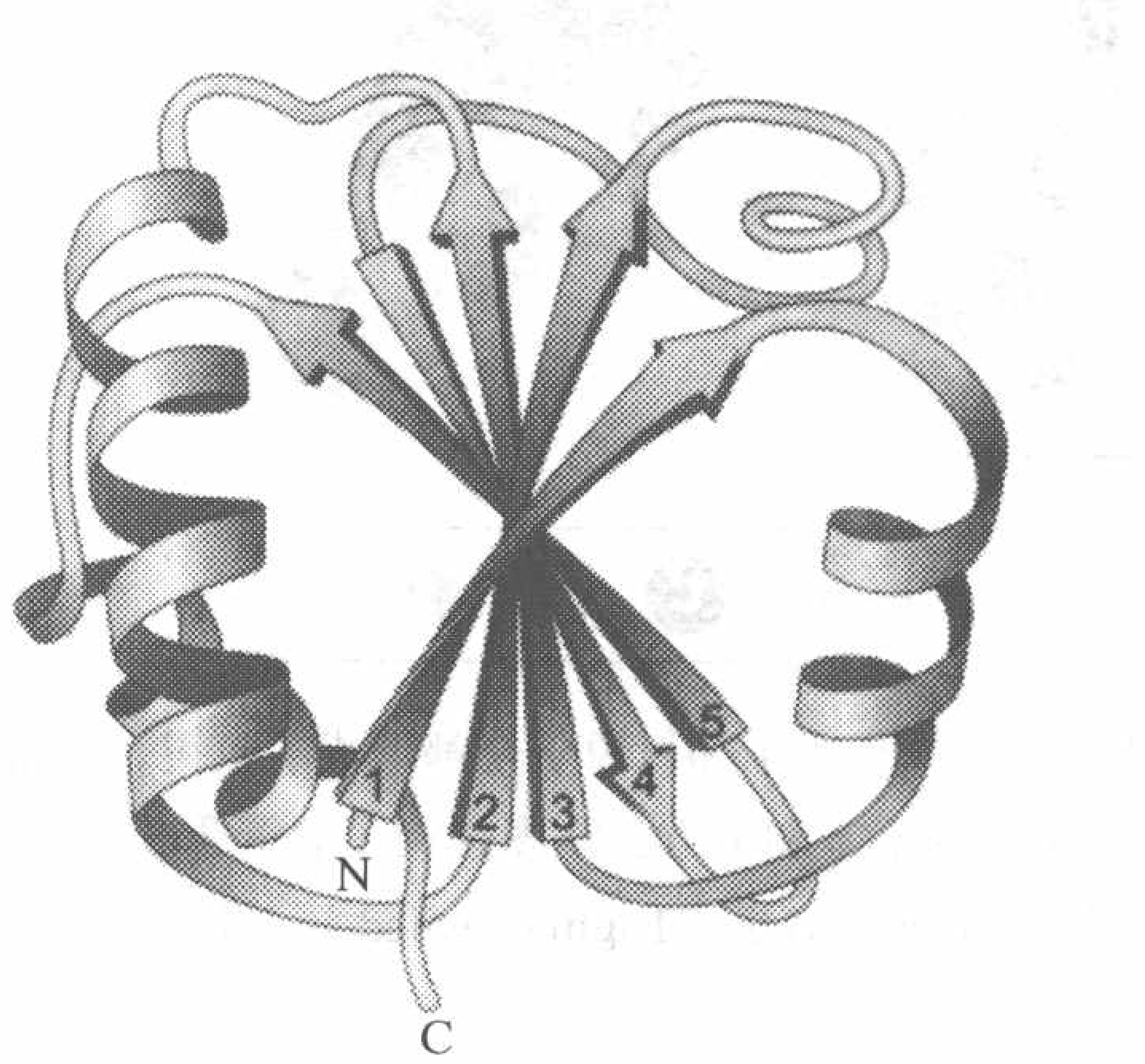


图 5-11 β 折叠沿着长轴形成右手缠绕。 图上展示了 *E. coli* 的硫氧还蛋白的混合结构。β 折叠单链以由蛋白质氨基端指向羧基端的箭头表示。（来源：Adapted from Branden C. and Tooze J. 1999. Introduction to protein structure，2nd edition，p. 20，Fig 2. 7a. Garland Science/Taylor & Francis LLC. © B. Furugren.）

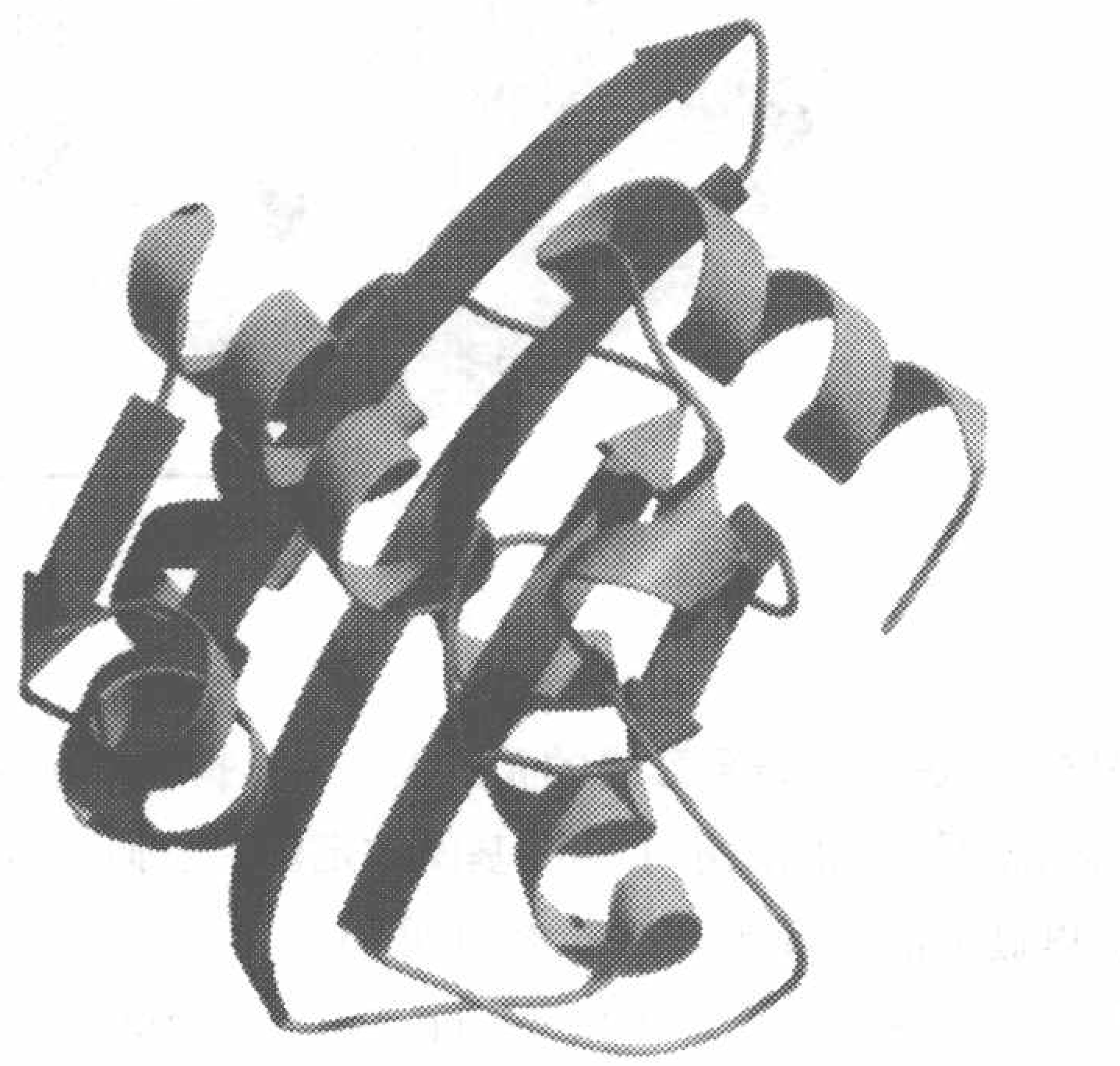


图 5-12 蛋白质结构的规则性和不规则性。 蛋白质骨架（绿色部分）的不规则构型使蛋白质可以与其他区域形成尽可能多的二级结构（紫色为 β 折叠，青色为 α 螺旋）。图中所示为腺病毒 E1 蛋白的结构。（Enemark E. J.，Chen G.，Vaughn D. E.，Stenlund A.，and Joshua-Tor L. 2000. Mol. Cell. 6：149.）镜像图用 MolScript、BobScript 和 Raster 3D 软件制作。

能形成最大数量的弱化学键，蛋白质不得不形成更多式样的外形。蛋白质多肽链的骨架趋向于形成α螺旋或β折叠，而侧链却趋向于缠绕骨架形成不规则构型，以此来使其形成的次要化学键达到最大强度。蛋白质多肽链的三维结构在两者之间采取了折中方案（图 5-12）。

正如我们将要在下文详细讨论的，对蛋白质折叠的最有力的影响因素之一可以归结于疏水性（非极性）氨基酸侧链包埋于蛋白质结构的核心部分。这使研究者们预测：在水溶液中，具有大量非极性侧链的蛋白质将会使其非极性残基内敛，从而比具有极性基团的蛋白质更加稳定。如果我们破坏一个由大量分子内氢键连接的极性分子，那么自由能的下降通常较小，因为极性基团可以转而与水分子形成氢键。相反，当我们破坏一个具有许多非极性基团的分子时，通常情况下自由能就会大量释放，因为这种破坏必然使非极性基团插入水分子。

氢键的排列方式决定蛋白质的特定构型

虽然稳定蛋白质的一部分能量来自于疏水相互作用，但是蛋白质结构的特定构型主要由氢键决定。伴随蛋白质疏水性稳定作用的能量没有方向性，而氢键需要准确的键长和键角（图 3-9 和表 3-3）。在一般情况下，蛋白质内部的所有氢键供体和受体都具有合适的配体。从能量上来说，蛋白质内部不能形成氢键是非常不利的，其数值为每个氢键几卡路里。蛋白质中氢键的重要作用在于，它不但使正确的结构稳定，而且使不正确的结构不稳定。

满足多肽骨架中所有氢键供体和受体（每个残基两个）的必要性导致了大多数蛋白质中巨大的α螺旋和β折叠区域的形成。一种多肽需要而且必须穿越某种蛋白质的非水溶性内部结构，以满足形成的氢键的条件，其唯一途径就是形成规则的二级结构。侧链没有足够的供体和受体来行使这一功能。所以，所有的大蛋白质分子都具有重要的β折叠区、α螺旋区或同时具有两者。除了数量较少的二级结构元件之外，能够由这些二级结构元件组合而成的蛋白质结构的种类几乎是无限的。即使是完全由β折叠或α螺旋组成的蛋白质也可以形成多种多样的结构（图 5-13）。

当然，某些肽段为了允许其在α螺旋和β折叠的单链（β单链）的末端转角（turn），必然变得不太规则。转角是氨基酸形成的圈，它可以连接α螺旋和β折叠，但在其单独存在的情况下不会表现出有意义的二级结构。转角的长度变化很大，可以从几个氨基酸到实际上更长的延伸肽段。然而，转角通常相对较短，这可以使伴随其形成而产生的未饱和氢键的数量达到最少（图 5-14）。

另外，这些圈的不规则结构在小分子的结合位点、酶的活性位点以及蛋白质-蛋白质相互作用的分子表面的形成中起着决定作用。正如我们在本章的剩余部分以及本书的其他章节将要讨论到的，这一点在蛋白质的三维结构中是显而易见的。

α螺旋相互结合形成卷曲螺旋

许多多肽链通过α螺旋相互间形成超螺旋而发生相互作用。典型的情况是，只有当一个α螺旋的非极性侧链的排列情况允许其与另一个螺旋连接，这样的相互作用才能存

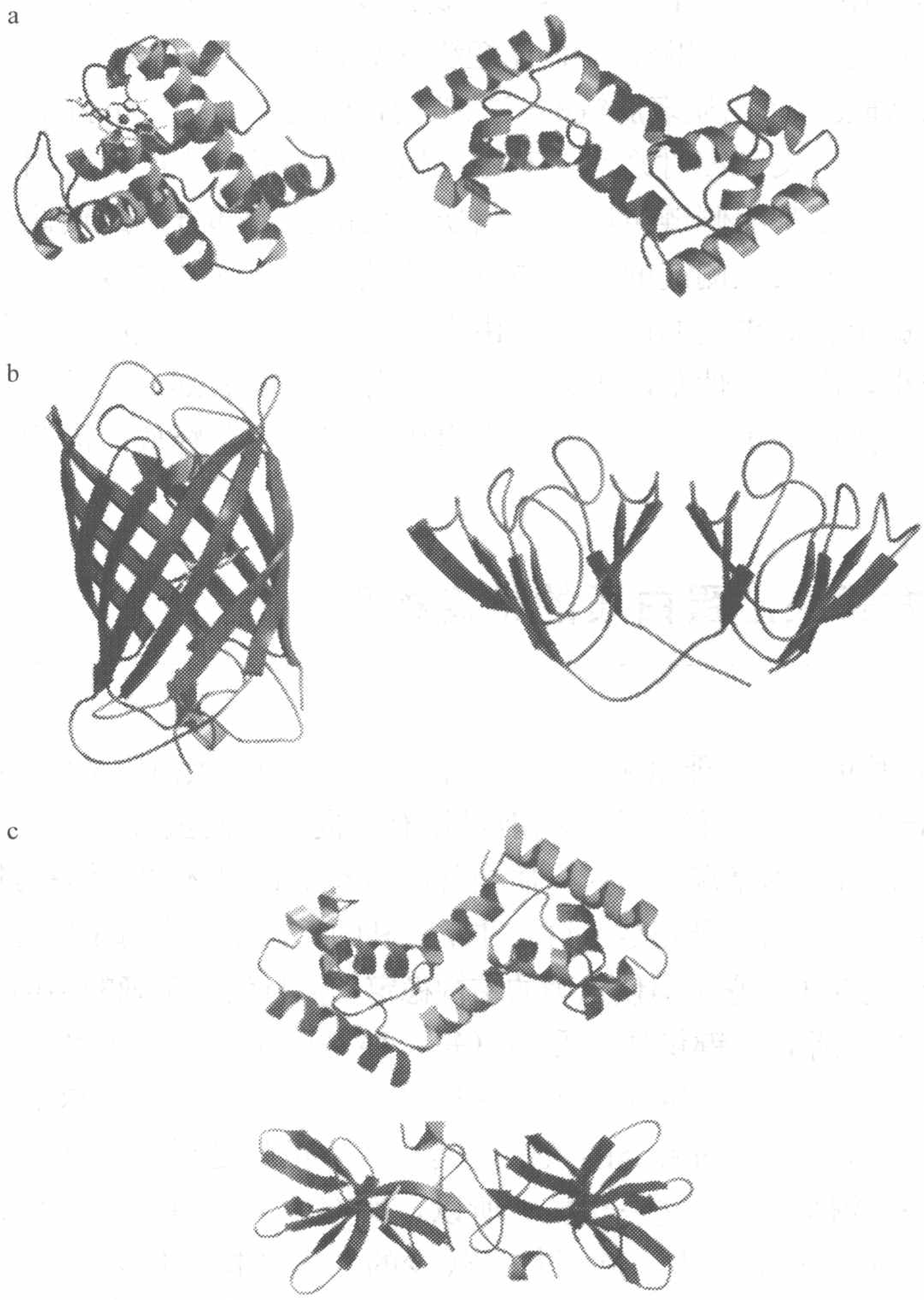


图 5-13　多肽链折叠。a. 含有 α 螺旋的蛋白质：肌球蛋白和 λ 受体的 N 端结构域；b. 含有 β 折叠的蛋白质：绿色荧光蛋白（GFP）和 γ crystalline；c. 含有 α 螺旋的 λ 受体的 N 端结构域与含有 β 折叠的 λ 受体的 C 端结构域的比较。（a. Vojtechovsky J.，Berendzen J.，Chu K.，Schlichting I.，and Sweet R. M. submitted and；Beamer L. J. and Pabo C. O.，1992，J. Mol. Biol. 227：177；b. Ormo M.，Cubitt A. B.，Kallio K.，Gross L. A.，Tsien R. Y.，and Remington S. J.，1996，Science 273：1392 and Chirgadze Y. N.，Driessen H. P. C.，Wright G.，Slingsby C.，Hay R. E.，and Lindley P. F.，1996，Acta Crystallographer D. Biol. Crytallogr. 52：712；c. Beamer L. J. and Pabo C. O.，1992，J. Mol. Biol. 227：177 and Bell C. E.，Frescura P.，and Hochschild A.，2000，Cell 101：801.）所有镜像图均由 MolScript、BobScript 和 Raster 3D 软件制作。

在。螺旋的相互缠绕影响了 α 螺旋的非整数（每圈 3.6 个残基）特性。这一特性使得只有当 α 螺旋以 18°角平行相互作用时，侧链基团才能整齐排列。如果 α 螺旋仍然保持完美的精确性，它们将只能与极少数残基连接。但是随着左手超螺旋的形成，整齐排列的高度稳定的卷曲螺旋（coiled-coil）就产生了（图 5-15）。

卷曲螺旋的一个具体例子是 DNA 结合蛋白中的亮氨酸拉链结构家族。这类 DNA 结合因子具有两个亚基，后者通过卷曲螺旋结构域相互结合形成二聚物。这种卷曲螺旋结

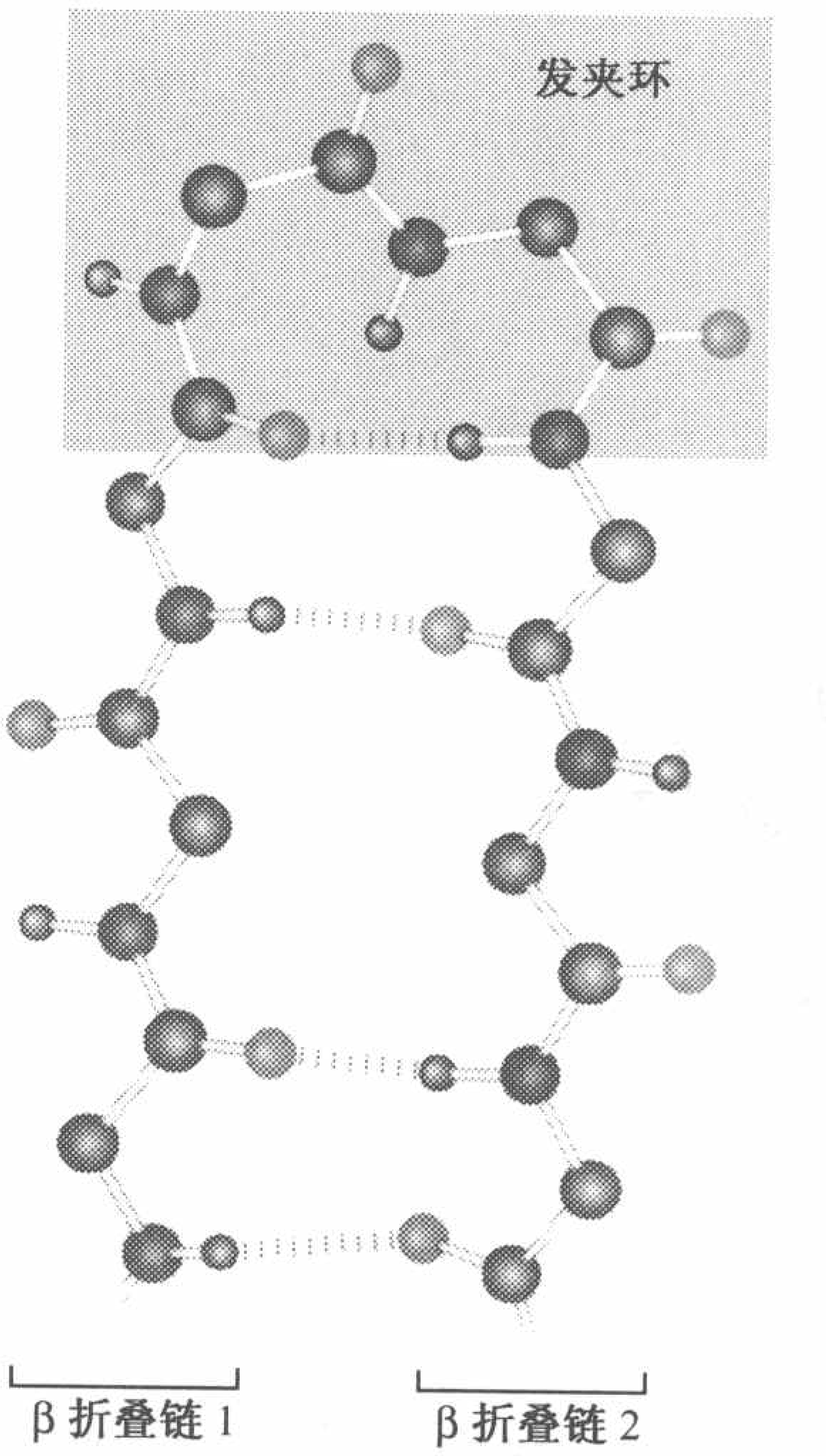


图 5-14　邻近的反平行 β 折叠中的发夹结构。图中展示的是一个由两个氨基酸残基构成的发夹结构的例子。发夹结构（结构顶端的阴影部分）中的化学键以绿色表示。

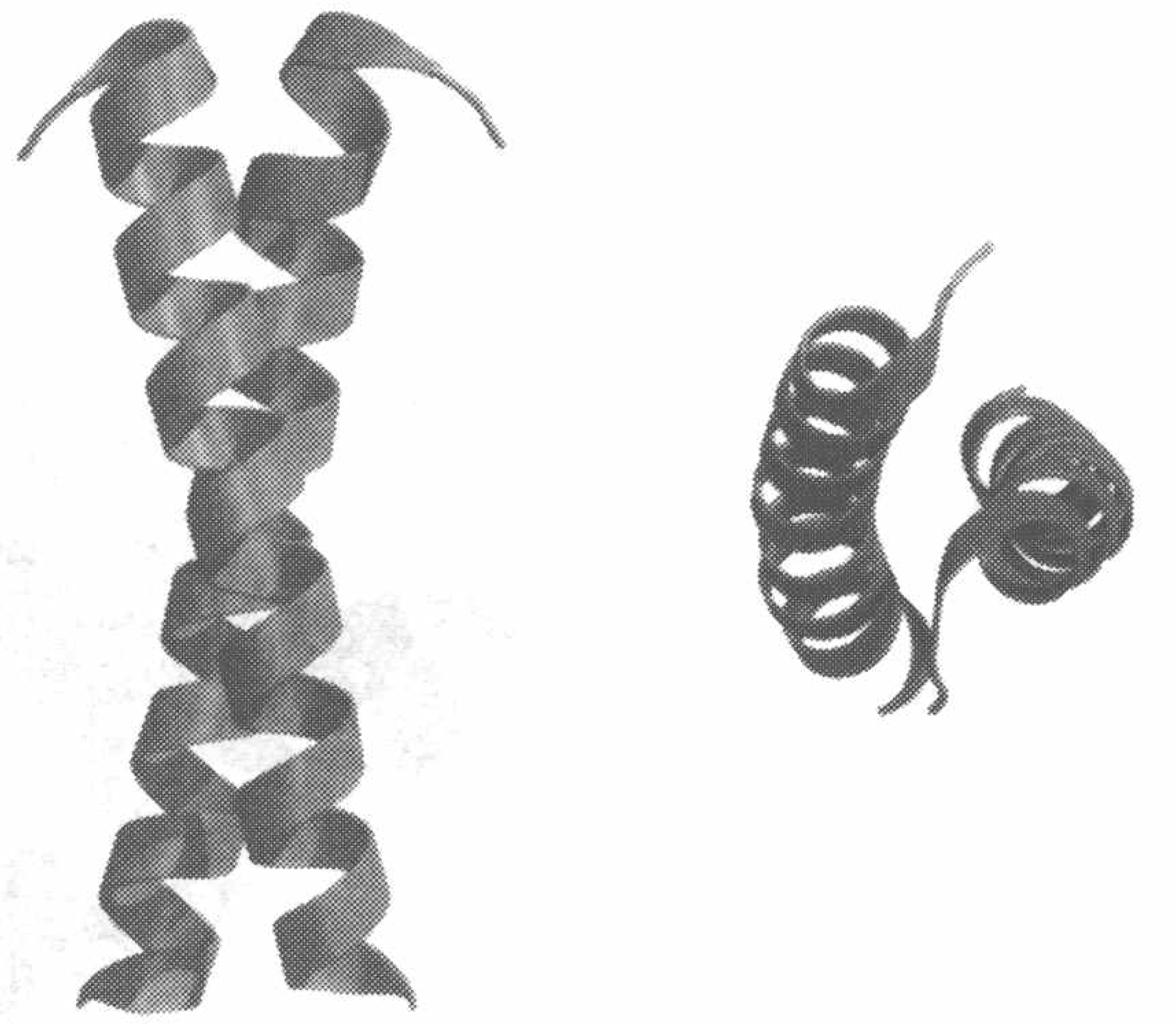


图 5-15　酵母转录因子 Gcn4 的亮氨酸拉链结构。亮氨酸拉链结构是盘绕再盘绕（见正文部分）的一个例子。在此，我们展示了亮氨酸拉链结构的两个视角：侧视（左图）和俯视（右图）。（Ellenberger T. E.，Brandl C. J.，Struhl K.，and Harrison S. C.，1992，Cell 71：1223.）镜像图由 MolScript、BobScript 和 Raster 3D 软件制作。

构域被称为亮氨酸拉链结构，这是因为在此结构中亮氨酸或其他具有脂肪族侧链基团的氨基酸，如缬氨酸或异亮氨酸，重复出现。这些亮氨酸以如下所示的规则模式出现。如果考虑一个 α 螺旋的两个转角，这将代表一个由大概 7 个氨基酸所组成的肽段。脂肪族氨基酸总是定位于这种肽段的第一位和第四位。这种定位确保 α 螺旋的一侧都是脂肪族的侧链，因为第一位和第四位都处于螺旋的同侧。两个邻近螺旋的这一侧面相互靠近，使得它们的疏水侧链远离水溶液环境。

大多数蛋白质由两个或三个结构域装配而成

可溶性蛋白质的亚基大小差别很大，其组成氨基酸残基数量可以从少于 100 个到多于 2000 个。形成折叠蛋白质的最小多肽的分子质量大约为 11kDa（大约 100 个残基），但是大多数单个亚基的分子质量介于 20～70kDa。

大于 20kDa 的蛋白质经常由两个或多个结构域组成（图 5-16；同时见框 5-2）。结构域（domain）习惯上被用来描述结构中与其他部分相对独立的那一部分，该部分看上去似乎能单独在溶液中保持稳定，实际情况也是如此。通常情况下，一个独立的结构域是由连续的氨基酸序列组成，而不是从多肽链中若干分离的部分组成。当考虑多结构域蛋白质是如何折叠时，这一点是非常重要的（图 13-28）。



图 5-16 丙酮酸激酶包含有邻近结构域。该酶的主要结构域分别以青色、紫色和红色表示。（Allen S. C. and Muirhead H. 1996. Acta Crystallogr. D. Biol. Crystallgr. 52：499.）镜像图由 MolScript、BobScript 和 Raster 3D 软件制作。

框 5-2 大蛋白质通常由数个小多肽链组成

大多数大蛋白质是数个小多肽链的规则集合体。多肽链间形成这样一个蛋白质分子的联系被称为蛋白质的四级结构。例如，负责 RNA 合成的大分子复合物（RNA 聚合酶）和蛋白质合成的大分子复合物（核糖体）都各自拥有多个亚基。这两种复合物都各自大概 500kDa 和 2500kDa，但并不具有任何大于 200kDa 的单独亚基。核糖体同时拥有蛋白质和 RNA 亚基，这类因子被称为核糖核蛋白（RNP）。

为什么大蛋白质复合物具有多个亚基而不是一个单独的大亚基？运用多个亚基组成大蛋白质复合物反映了一个建筑原则，这一原则适用于所有复杂结构，不论是无生命的还是有生命的。这一原则说明，如果有缺陷的亚基能够在被混入最终产物内之前被丢弃，我们就能更容易地减少结构错误的冲击。例如，让我们考虑以两种可选方案用 1000 000 个原子组成一个分子。在方案 1 中，我们一个原子接一个原子的构建分子；在方案 2 中，我们首先构建 1000 个每个含有 1000 个原子的小的单元，然后再用这些单元组成含有 1000 000 个原子的最终产物。现在考虑这样一个情况，假设构建过程会发生随机错误，每 100 000 次出现一次错误的原子插入。我们可以假设任何一次错误都会导致产物的无功能。

在方案 1 中，每个分子将会拥有平均 10 个错误原子，所以几乎没有合成出正常的产物。然而在方案 2 中，仅仅 1% 的亚基存在错误。如果有去除错误亚基的机制存在，正常的产物能够很容易产生，而细胞也将不会被不时出现无功能的亚基所困扰。这同样也是构成装配流水线理论基础的构建战略，在流水线上复杂的工业产品，如无线电和汽

车，被生产出来。在装配过程中的每一步都有去除错误元件的机制。起初在工业装配流水线上，错误是由人工去除的；现在，自动化已经取代了人工控制。在细胞中，错误有时是由特定的酶控制的。如果一个亚基单体被错误地装配，它通常不能被聚合酶所识别，从而不会被混入大分子。在另一种情况中，错误亚基由于不能自发成为稳定的分子集合体的一部分而被去除。

蛋白质由少数几种结构模式所组成

开始确定了大约半打蛋白质的结构，表明蛋白质的折叠模式令人眼花缭乱，以此推测必然存在无限多种蛋白质结构。现在我们知道了上千种蛋白质的三维结构，然而多种多样的蛋白质结构大部分是由数目相对较少的结构域构成。尽管确切的数量无法估计，但实际存在的独立的结构域模式的数量要比独立蛋白质的数量小若干个数量级。

特定种类的结构域模式经常是与特定种类的活性联系在一起的。一个常见的模式被称为双核苷酸折叠（dinucleotide fold），因为这一结构经常在与 ATP 结合的酶中被发现（图 5-17）。这种结构域具有处于中心部分的平行 β 折叠和处于两侧的 α 螺旋，并通过该结构结合 ATP。核苷酸结合位点在 β 折叠的 C 端。不同之处在于 α 螺旋的数量和具体排列方式以及稍加补充的 β 折叠的次序。具有相似结构的同源结构域在许多不同蛋白质中形成相同的功能。

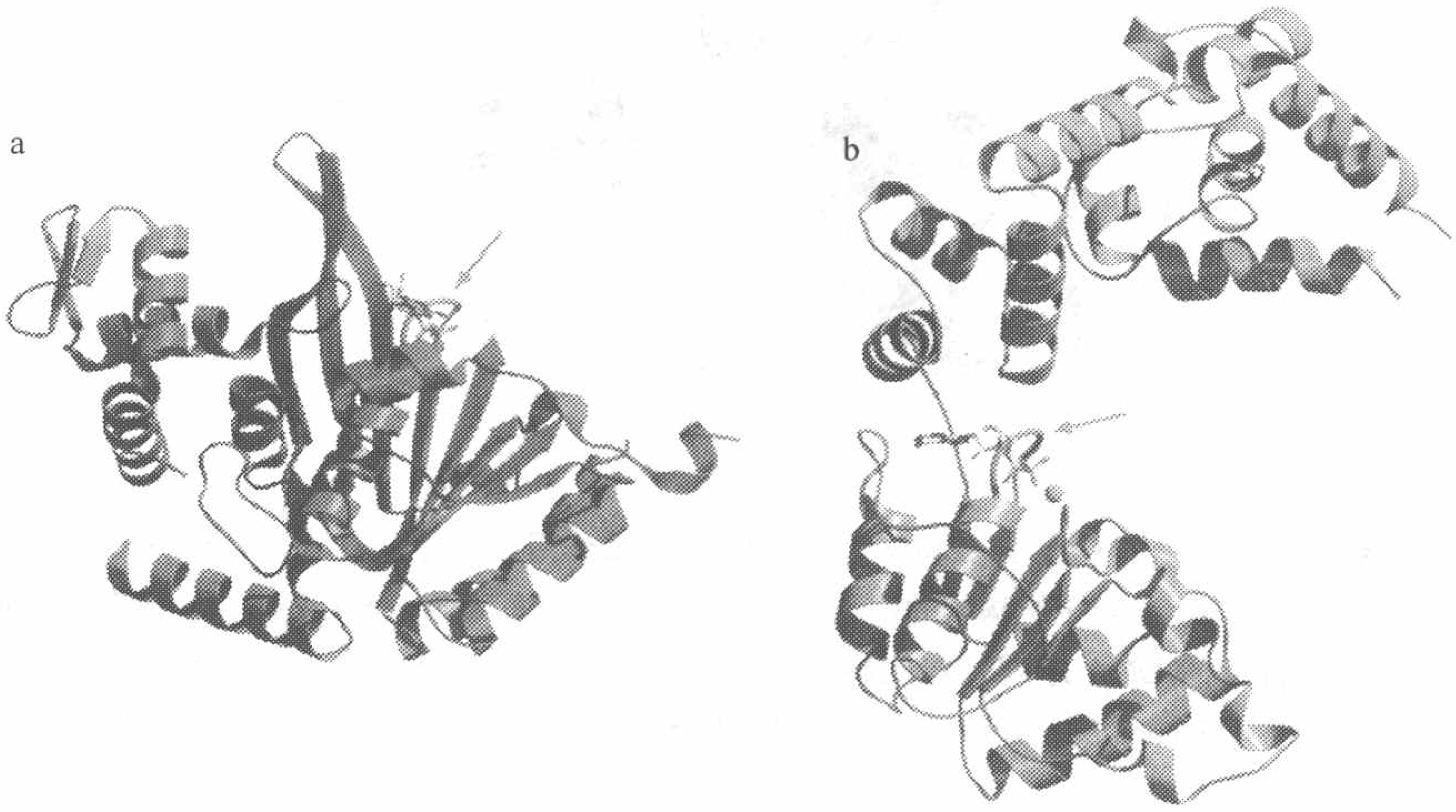


图 5-17 结合 ATP 的酶。红色箭头指出结合于各自结构中的 ATP 分子。a. RecA；b. DnaA。（a. Story R. M. and Steitz T. A. 1992，Nature 355：374；b. Erzberger J. P.，Pirruccello M. M.，and Berger J. M.，2002，EMBO J. 21：4763-4773.）镜像图由 MolScript、BobScript 和 Raster 3D 软件制作。

蛋白质的不同功能来自于结构域组合的多样性

不同规格的计算机能够由合适的组成模块装配而成，与此类似，蛋白质功能的多样性似乎来自于其模块化的构成。在此我们可以给出众多的实例。例如，自然界存在许多

对应于不同特定底物的脱氢酶。每种酶都具有两个结构域，一个是普通的结合辅酶 NAD^+ 的双核苷酸结合结构域，另一个是具有催化位点的底物结合结构域。在不同脱氢酶中，后者的结构有很大变化。

调控基因表达的阻遏蛋白和活化蛋白是模块化构成的另一个实例。*E. coli* 的 Lac 阻遏蛋白和新陈代谢副产物基因活化蛋白（第 16 章）（CAP）都具有多个结构域。CAP 的结晶结构表明它具有两个结构域：较大的结构域在其内部结合一分子的环状 AMP，而较小的结构域识别特定的 DNA 序列（图 5-18）。在 CAP 的 cAMP 结合结构域和依赖 cAMP 的蛋白激酶的调控亚基之间存在着明显的氨基酸序列的相似性，这表明两种蛋白质的 cAMP 结合结构域来自相同前体。在 CAP 中，cAMP 结合结构域与 DNA 结合结构域相连，因此 cAMP 水平的变化调控着转录水平。在激酶中，cAMP 结合结构域调控着参与酶联放大反应的一连串酶中第一个酶的活性，这一反应导致了体内储存糖原的裂解。

在 Lac 阻遏蛋白这一实例中，该蛋白质还拥有一个 DNA 结合结构域和另一个结合乳糖的结构域。该 Lac 阻遏蛋白的 DNA 结合结构域与 CAP 十分相似，然而其相连的另一结构域——Lac 阻遏蛋白的乳糖结合结构域和 CAP 的 cAMP 结合结构域，却有很大的不同。

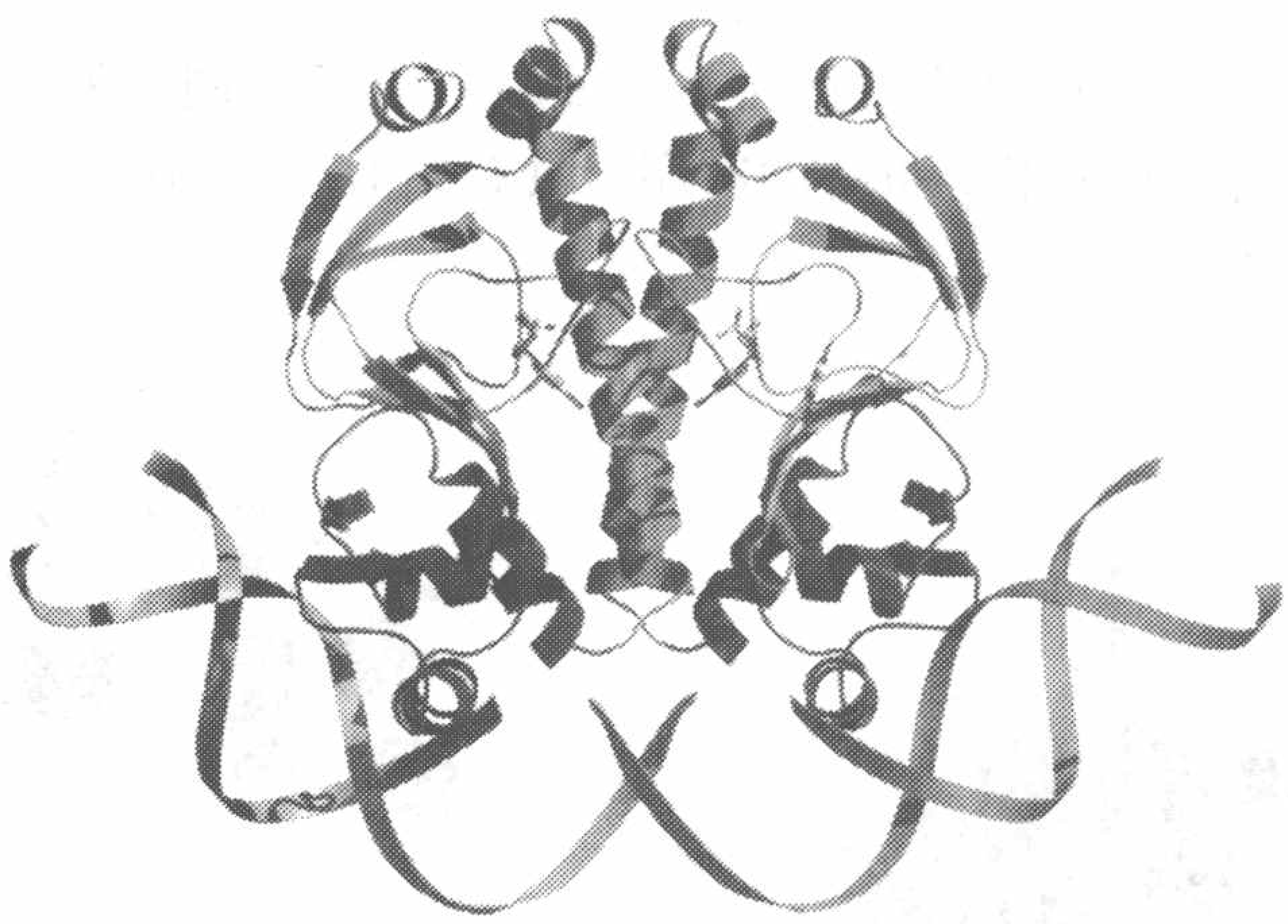


图 5-18 结合有 cAMP 的 CAP 复合物与弯曲的 DNA 相互作用。以青色表示的 CAP 中较大的结构域与环状 AMP 结合，后者处于结构域的中心部分，以红色和黄色表示。较小的 DNA 结合结构域（紫色部分）识别特定的 DNA 序列（双螺旋以红色和灰色表示）。（Schultz S. C. Shields G. C. , and Steitz T. A. 1999，Science 253：1001. ）镜像图由 MolScript、BobScript 和 Raster 3D 软件制作。

弱化学键决定蛋白质在 DNA 和 RNA 分子上的正确位置

DNA 结合蛋白介导了生物学中许多重要的生物过程。将这些蛋白质与 DNA 连接在一起的化学键仍然是那几类弱化学键，正是它们使蛋白质、DNA 和 RNA 拥有了各自独特的三维结构。占 DNA 结合蛋白最大部分的那类蛋白质在包裹和压紧大量 DNA 的过程中起着结构性的作用，而这一过程对细胞来说是必需的。例如，人类细胞的细胞

核直径仅有10μm（10^{-5}m）却包含有几乎2m长的双链DNA（第7章）。

蛋白质有许多方法来识别DNA。某些蛋白质-DNA相互作用只对特定的DNA序列起作用，而另一些却对处于特定构型的DNA更具选择性。例如，在DNA复制或重组过程中，如果细胞内的DNA没有受伤，单链DNA会迅速被单链DNA结合蛋白（SSB）结合。这些蛋白质的结合不太依赖于DNA序列的特异性，而是对单链和双链DNA的区别具有高度选择性。为了达到这种选择性，SSB和单链DNA间的主要的相互作用是磷酸盐骨架上的离子键或氢键相互作用，或者是碱基间庞大的环状侧链（如Tyr或Trp）的插入作用（图5-19）。

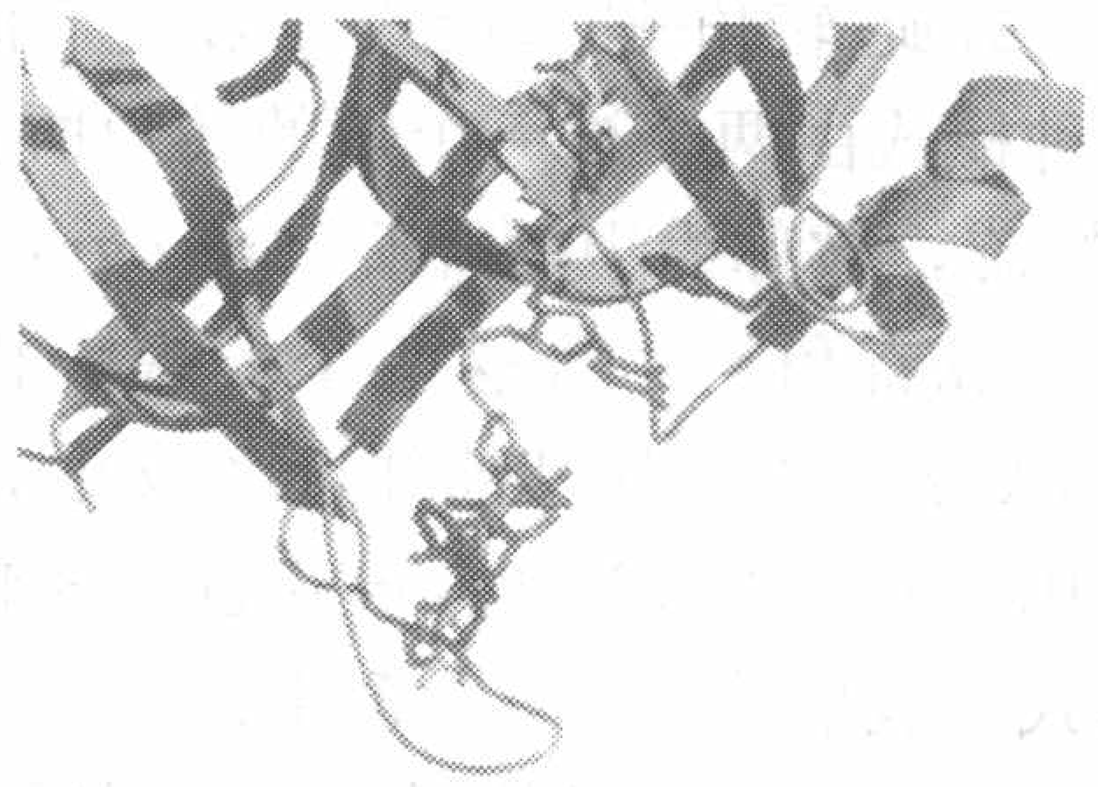


图5-19　单链DNA结合蛋白（SSB）和单链DNA相互作用。SSB以灰色表示，而单链DNA以红色表示。（Raghumathan S.，Kozlov A. G.，Lohman T. M.，and Waksman G. 2000，Nature Structural Biology，8：648.）镜像图由MolScript、BobScript和Raster 3D软件制作。

我们将要在本书中探讨的大多数DNA结合蛋白识别双链DNA中的特定DNA序列。这类蛋白质经常参与选择基因组中的特定序列作为转录或复制的起始位点，或者参与其他DNA的加工。实际上，2%～3%的原核细胞蛋白质和6%～7%的真核细胞蛋白质已知是，或者被预测是序列特异性的DNA结合蛋白。蛋白质识别特定DNA序列的最通常的机制为α螺旋对DNA中所谓的大沟的插入（图5-20）。正如在图5-2中显示的以及在图6-1中明确展示的，双螺旋具有一个宽的沟，即大沟，和一个窄的小沟。将α螺旋插入大沟以此来识别是有利的，有以下几个原因。

1. 大沟的宽度和深度对α螺旋的尺寸都非常适合，使DNA和大约半个α螺旋弱相互作用得以形成。

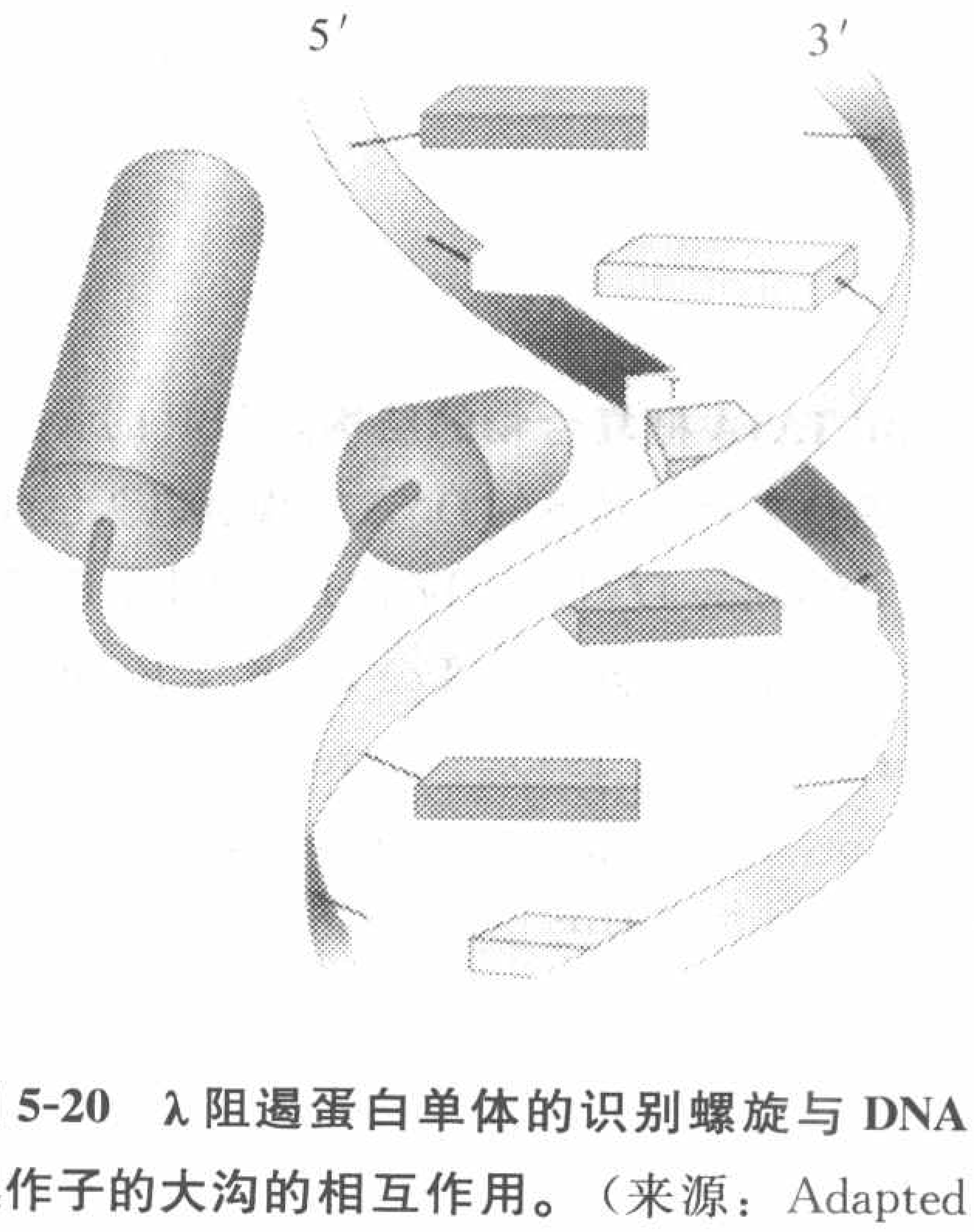


图5-20　λ阻遏蛋白单体的识别螺旋与DNA操作子的大沟的相互作用。（来源：Adapted from Jordan S. R.，and Pabo C. O.，1988，Structure of the lambda complex at 2.5 Å resolution. Science 242：893-899. © AAAS.）

2. 大沟富含位于碱基边缘的氢键受体和供体（图6-10）。更重要的是，对每个碱基对来说，氢键元件的模式都各不相同。这使得氢键供体和受体的模式可以作为DNA序列的密码，运用同样的方法，碱基对间的氢键保证了DNA杂交过程中互补DNA序列的正确识别。一幅大沟中每一对碱基对的氢键供体和受体残基的模式的图清楚地表明了每对碱基对的模式区别（图6-10）。注意，并非仅是G∶C碱基对能够容易地与A∶T碱基对相区分，A∶T与T∶A和G∶C与C∶G同样能够被区分。相反，小沟中的碱基对模式只能提供很少的信息，通常只能将G∶C与A∶T分开。

3. α螺旋具有偶极子，这使其N端带有正电荷。这一正电荷末端经常与靠近大沟的磷酸盐骨架产生弱相互作用。

螺旋-转角-螺旋模式是第一个被确认的参与序列特异的 DNA 结合的蛋白质模式。这种模式由两个被一个短的转角所分隔的相邻 α 螺旋所组成（图 5-20 和图 5-21）。一个 α 螺旋，被称为识别螺旋（recognition helix），负责 DNA 的识别；另一个 α 螺旋与前者大概成直角。尽管这两个 α 螺旋构成了 DNA 识别模式的核心部分，螺旋-转角-螺旋 DNA 结合蛋白的其他邻近区域通常起到稳定这两个 α 螺旋的排列方式和结合 DNA 的作用。其他 DNA 结合模式也将 α 螺旋插入到大沟中，如锌指和亮氨酸拉链 DNA 结合模式（我们将在第 17 章讨论）。

虽然 α 螺旋是特异性的识别 DNA 的优势型，而有些蛋白质运用的则是不同的策略。其中一个极端的例子在 TATA 结合蛋白（TBP）中被发现。在许多真核启动子中该蛋白质决定了转录起始的位点（第 12 章）。TBP 运用 β 折叠的一个延伸区域识别所谓的 TATA 框的小沟（图 5-22）。因此，在本例中，我们看到了 β 折叠的使用而非 α 螺旋，看到了与小沟的相互作用而非大沟（详情请见第 12 章）。

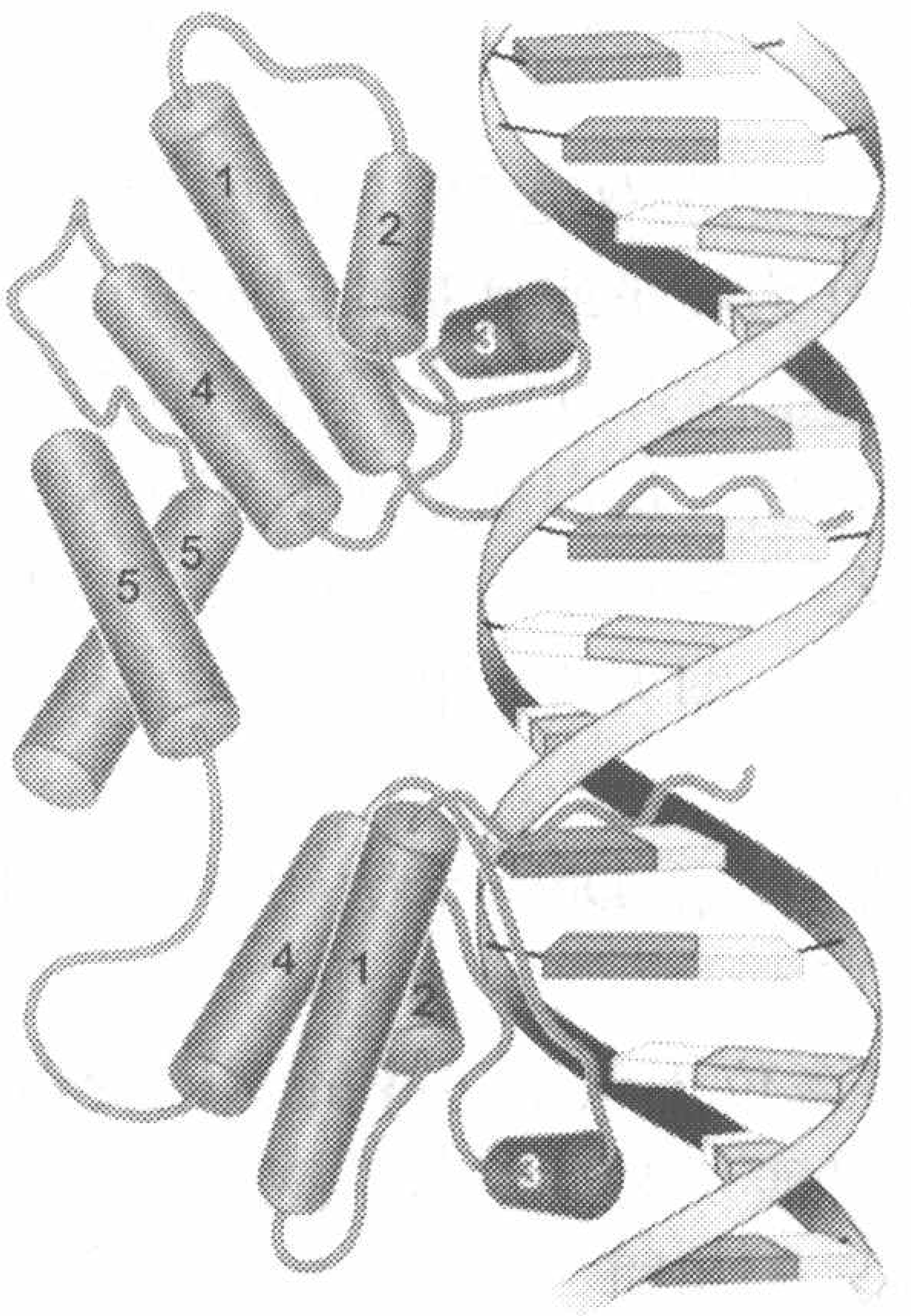


图 5-21　λ 阻遏蛋白-操作子复合物的几何构型。图中展示了两个与操作子结合的 λ 阻抑物单体，每一单体中的螺旋都以 1～5 标记。正如在图 5-20 中展示的那样，螺旋 3 插入到了大沟中。（来源：Adapted from Jordan S. R., and Pabo C. O., 1988. Structure of the lambda complex at 2.5 Å resolution. Science 242: 893-899, Fig. 2b. © AAAS.）

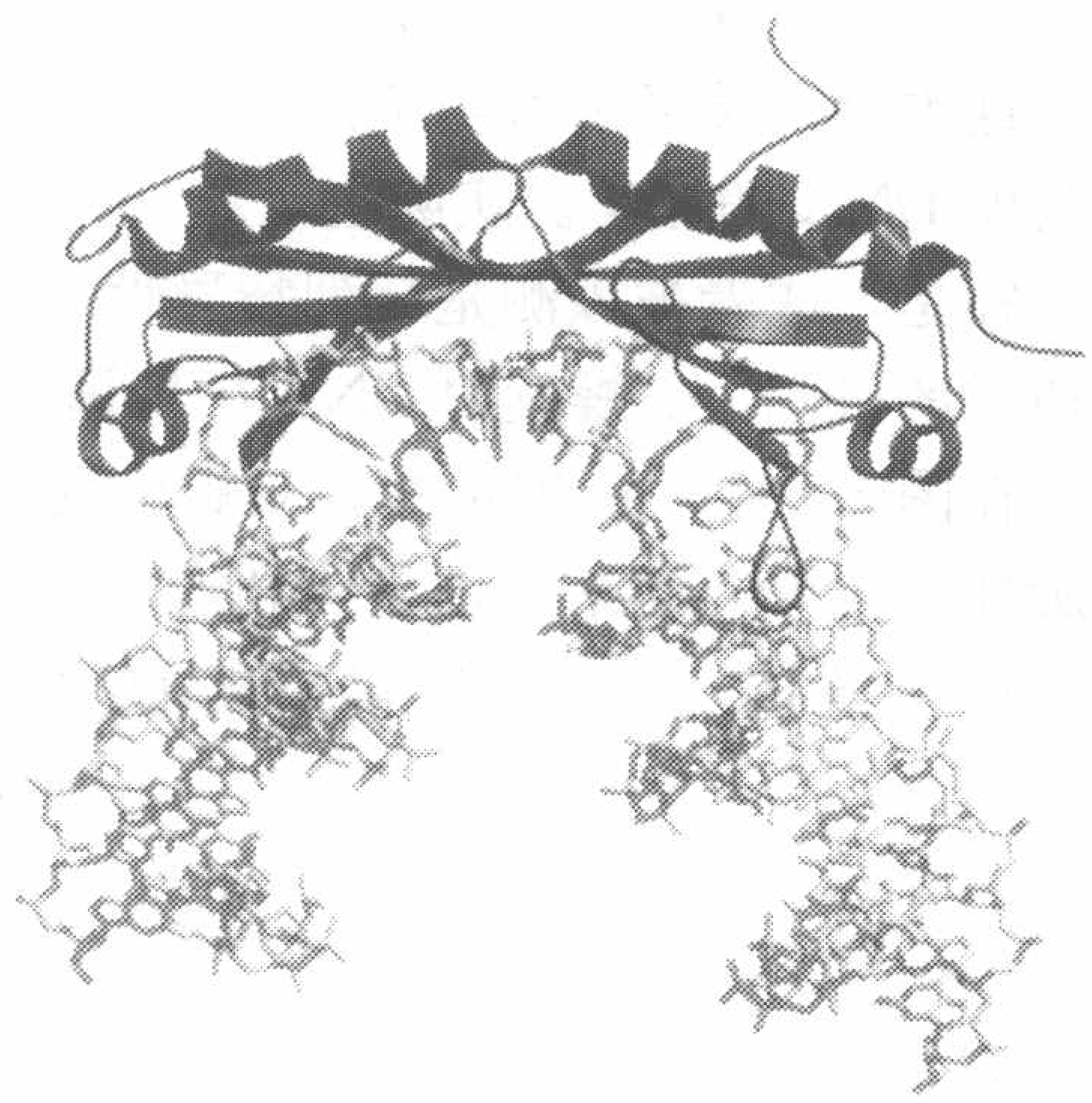


图 5-22　TBP-TATA 框复合物的结构。TBP 的骨架在图的上方以紫色表示；其下的 DNA 双螺旋以灰色和玫瑰色表示。（Nikolov D. B., Chen H., Halay E. D., Usheva A. A., Hisatake K., Lee D. K., Roeder R. G., and Burley S. K., 1995. Nature 377: 119.）镜像图由 MolScript, BobScript 和 Raster 3D 软件制作，图中两侧延伸的 DNA 模型由 Leemor Joshua-Tor 制作。

蛋白质沿 DNA 扫描寻找特定的 DNA 结合位点

许多 DNA 结合蛋白除了与它们各自的识别位点的特定碱基对结合以外，还可以与 DNA 骨架形成牢固的结合。介导这种骨架结合的是那些带正电的氨基酸补丁，它们的

定位非常接近结合碱基对的氨基酸。这种连接主要依赖于这些正电性补丁和负电性的DNA磷酸盐骨架间的静电吸引力。由于磷酸盐骨架具有与序列无关的相似的负电性表面，所以这种蛋白质-DNA骨架结合同时促成了蛋白质对DNA的特异的以及非特异的牢固亲和性。因此，即使是高度特异性的DNA结合蛋白也会与非特异性的DNA位点具有牢固的亲和性。

例如，某些已经详细研究的基因表达调控子（如前面提到的乳糖阻遏蛋白）对其识别序列的亲和性比它们与非特异性DNA的亲和性要提高大约10^5倍。所以，在细胞中这类蛋白质除了结合于各自特异的靶序列以外还与相当数量的非特异性位点结合。这是因为与特异性位点相比，非特异性位点的数量要大的多。实际上，基因组中的每个核苷酸都能被作为一个潜在的（几乎总是非特异性的）结合位点。所以，在*E. coli*中将会有大约5×10^6个非特异性的结合位点，因为它的环状基因组有大约5×10^6bp。因此，即使特异性与非特异性的DNA结合亲和性的比率很高（10^5倍），非特异性与特异性的结合位点的比率更高（5×10^6倍）。这一比较解释了为何细胞不得不具有多个阻遏蛋白的拷贝，以保证对特异的调控DNA结合位点的持续占据。在这种情况下，大多数的阻遏蛋白分子都与非特异性位点结合了。

非特异性的蛋白质-DNA相互作用不仅仅是在DNA识别过程中蛋白质运用DNA骨架带电性质的必然结果。人们确信这种相互作用提高了一个给特定调控蛋白寻找其正确靶点的速度。由于非特异性结合蛋白质的静电相互作用，它们被迫沿着DNA线性迁移，而不是简单的在DNA上跳跃。这种迁移允许DNA结合蛋白在“寻找”特异性结合位点的过程中随机搜索结合位点。由于被限定为线性运动，蛋白质会比它们在细胞中自由迁移更快的速度抵达靶位点。

一小部分DNA结合蛋白并不是在DNA上简单的扩散，相反，它是沿着DNA主动地移动。在DNA的复制、修复和重组过程中，这些蛋白质通过沿着DNA的方向性的移动完成重要的功能（第8、9、10章）。因为这种运动是有方向性的，它需要能量供给。所以，这类蛋白质水解ATP来指导其DNA结合过程中的变化。

蛋白质识别RNA的不同策略

如前所述，RNA与DNA相比具有更多的结构类型。RNA结合蛋白在RNA的功能中扮演着多种角色，从RNA的稳定到酶催化加工。几种与其靶分子结合的RNA结合蛋白的结构显示它们具有不同的蛋白质-RNA识别策略。

某些RNA结合蛋白特异性地与双链RNA相互作用。在这类例子中，蛋白质识别那些将RNA与DNA双螺旋相区分的特征。例如，2′-羟基基团的存在就是RNA的一个明显的区分特征。与此类似的特征还有，RNA大多形成A型螺旋（第6章），这种螺旋具有比B型螺旋更深更窄的沟。与上文讨论的DNA结合蛋白相反，这类蛋白质并不是通过将α螺旋区域插入到RNA的沟中来识别核酸的。

许多重要的与RNA分子结合的RNA结合蛋白并非处于规则的螺旋构型。这些蛋白质也包括在转录和RNA加工过程中与信使RNA分子相互作用的蛋白质。同样地，剪接和翻译RNA的机器具有由RNA和蛋白质复合体而成的亚基（第13章和第14章）。

RNA 识别结构域（RNA-recognition motif，RRM，也称为核糖核蛋白，ribonuclear protein motif，RNP motif）是与 RNA 特异性识别有关的最常见的蛋白结构域，RRM 由 80～90 个氨基酸组成，形成一个四链的反平行 β 片层（sheet）和与之方向相反的两个 α 螺旋。该种排列使得该结构域具有变换 αβ 拓扑结构的特性。该种结构域的一个常见例子是 U1A 蛋白。该蛋白质与 U1 小核 RNA（snRNA）相互作用，同时剪切 RNA 转录体（第 13 章）；U1A：U1ASnRNA 复合物的结构如图 5-23 所示。U1A 的 RNA 结合表面的形状对该种 RNA 是特定的。

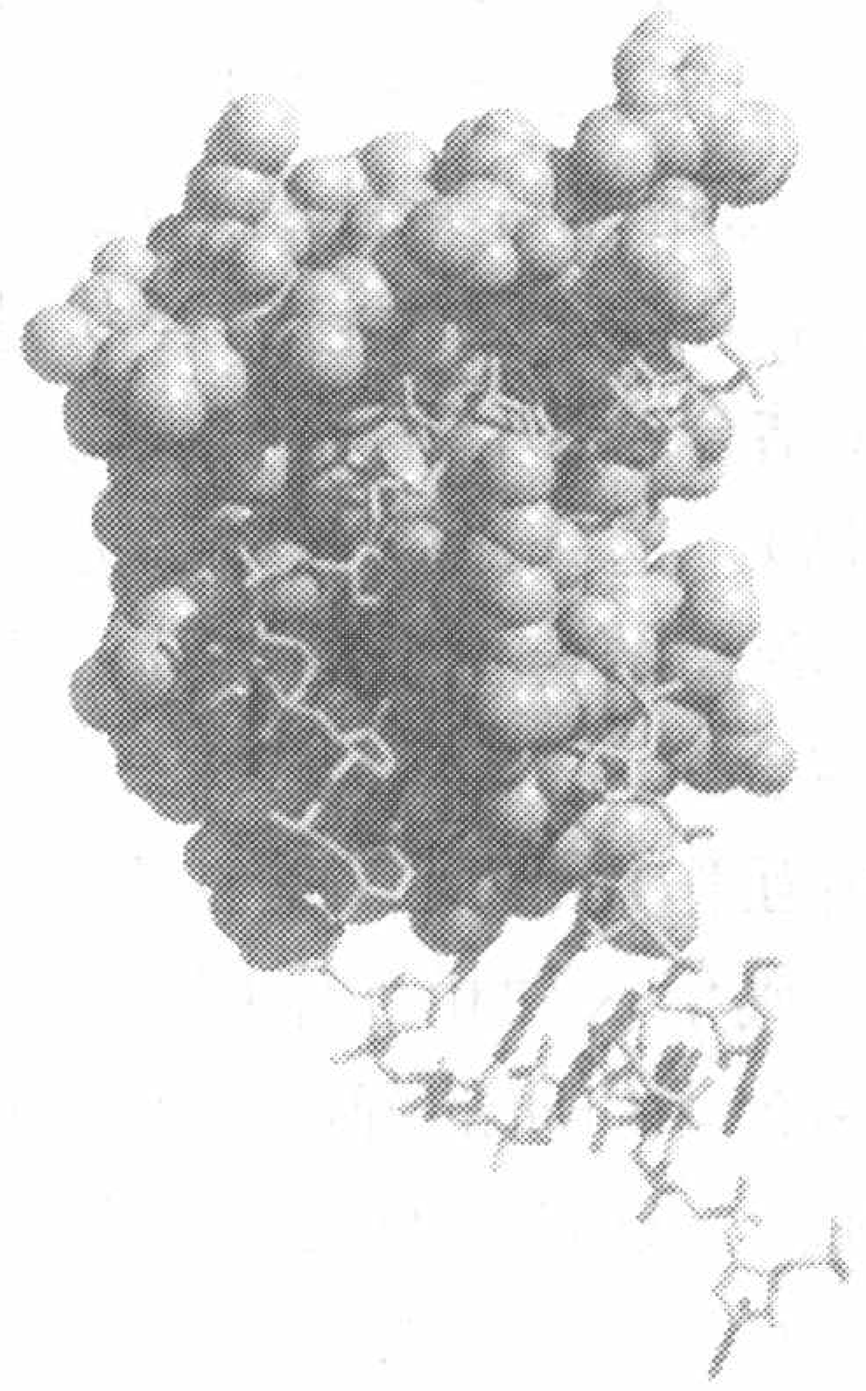


图 5-23　剪接体蛋白质-RNA 复合物的结构：U1A 结合于 U1 snRNA 的Ⅱ型发夹结构。蛋白质以灰色表示，U1 snRNA 以绿色表示。(Oubridge C.，Ito N.，Evans P. R.，Teo C. H.，and Nagai K. 1994. Nature 372：432.）镜像图由 MolScript、BobScript 和 Raster 3D 软件制作。

现在多种 RRM 的结构是已知的。大多数情况下，β 折叠介导了 RNA 的识别。尤其是其中的三个氨基酸残基与 RNA 特定接触：一个精氨酸或赖氨酸与磷酸骨架形成盐桥，两个芳香族氨基酸残基与核苷酸形成紧密相互作用。

单独的 RRM 能够识别相对较短的 RNA 区域（通常短至 4 个，最长 8 个核苷酸）。因此，通常识别 RNA 的蛋白质运用多个 RNA 识别结构域实现功能。如图所示，U1A 仅仅是两个 RRM 结构域的其中之一。在一些蛋白质中，三到四个结构域会共同作用以增强 RNA 结合的亲和性与特异性。除 RRM 外，还存在另外一些 RNA 识别结构域。因此，图 5-24 展示了一系列包含 RNA 识别结构域的蛋白质。

正如我们已经提到的，RNA 识别对于转录和剪切十分重要；但对于基因表达和 RNA 进程的其他方面同样至关重要。例如第 18 章将要讨论的基因表达的调控。这部分是通过 RNA 自身实现的，比如 microRNA 和 siRNA 介导的 RNA 干扰（RNAi）。我们将具体讨论通过 Dicer 蛋白调控的 RNA 识别和通过 PAZ 结构域调控的 Argonaut 蛋白（见图 5-24）。

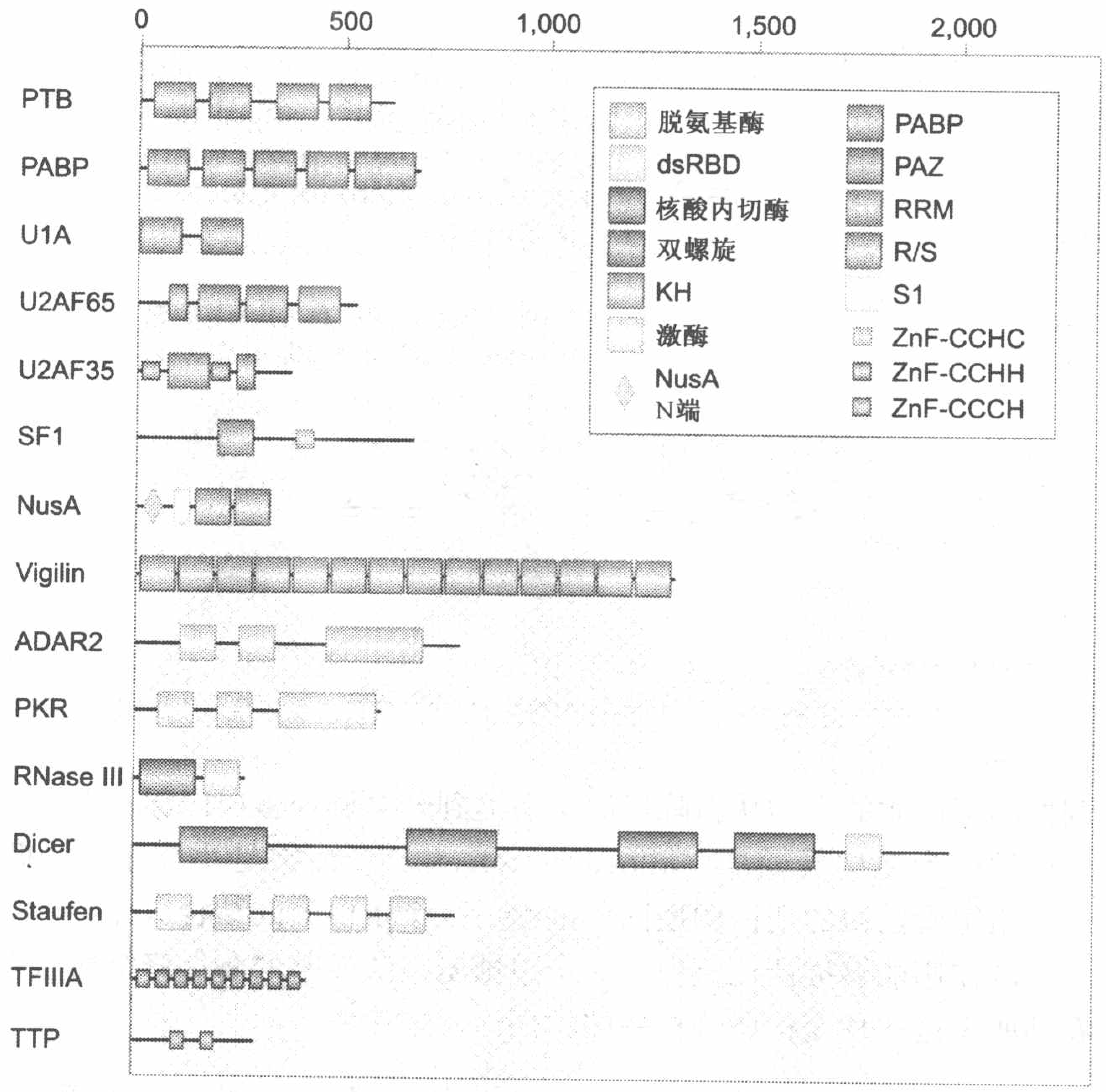


图 5-24　蛋白质中的 RNA 结合结构域。一些 RNA 结合蛋白家族的实例。如图所示，不同蛋白的结构域的类型、数目和排列均有所不同。图中还标出了其他功能结构域。许多该类蛋白将在后续章节再次提及。如转录（TFⅢA、PABP 和 NusA）、剪切（U1A 和 U2A）以及 RNR 调控（Dicer）。不同结构域以彩色方框显示：RNA 调节结构域（RRM，最常见的结构域）、K 同源结构域（KH，能够结合单链 RNA 和 DNA）、dsRBD（不依赖序列的 dsRNA 结合结构域）以及锌指结构域（ZnF）。酶活结构域以及一些不常见的结构域也一同列出。PABP，poly（A）结合结构域；PTB，结合 Polypyrimidine-tract；R/S，富含精氨酸和丝氨酸的结构域；SF1，剪切因子-1，TTP，tristeraprolin；U2AF，U2 auxilliary 因子（据 Lunde B M et al. 2007 Nat. Rev. Mol，Biol. 8：479，Fig. 1. © Macmillan，经许可重绘）

变构：通过改变蛋白质的外形调节其功能

不论是小分子还是大分子（配体），与蛋白质的结合都会引起其构型的实质性改变。这类配体引起的构型改变能够具有多种效果，也许是增加该蛋白质对第二个配体的亲和性，也许是将蛋白质的酶学活性打开或关闭。这就是变构调控（allosteric regulation），是生物系统中的普遍调控机制。“变构”的意思是“其他外形”，其基本机制如下所述：结合于蛋白质某一位点的配体改变了这个蛋白质的外形，作为这种改变的结果，该蛋白质其他部位的活性位点，或另一个结合位点，被改变以增加或减小其活性（图 5-25）。

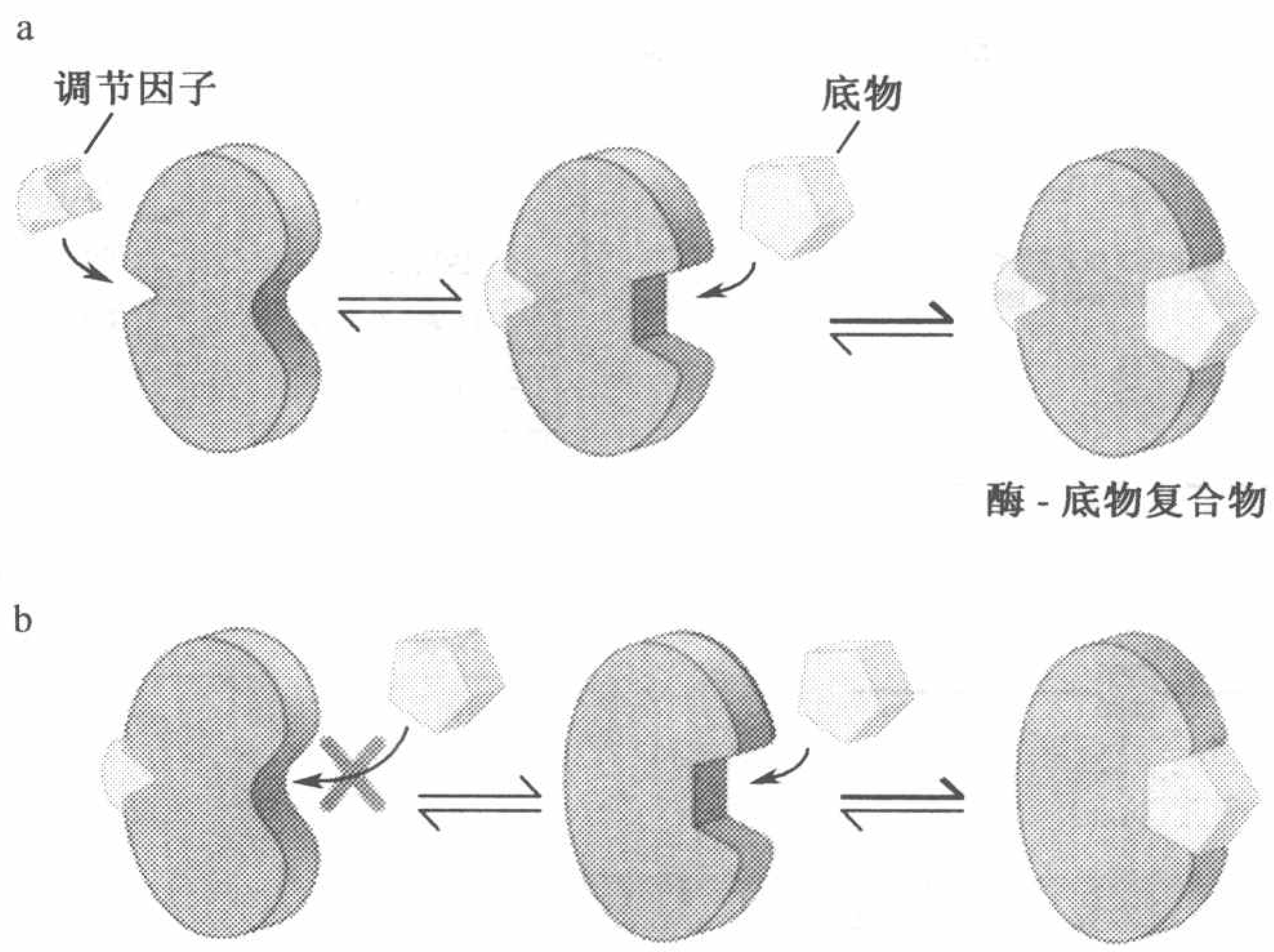


图 5-25 酶活性的变构调节。(a) 酶的正调控。调节因子的结合引起酶形状的改变，从而可见结合底物。(b) 酶的负调控。这种情况下，调节因子的结合锁定了酶的形状，从而不能结合底物。

受到如此调控的蛋白质的实例从新陈代谢的酶类到转录调控蛋白广泛存在。

配体，也被称为变构受动器（allosteric effector），经常是小分子——糖类或氨基酸。但是一个给定蛋白质的变构调控同样能由另一个蛋白质的结合所介导，在某些情况下，极为相似的作用能够被调控蛋白中一个单独氨基酸残基的酶促修饰所激发。我们将在本节中看到所有这 3 种变构调节机制的举例。

通过小配体、蛋白质-蛋白质相互作用和蛋白质修饰等例子了解变构调控的结构基础

在此，我们具体考虑 3 种变构调控机制的结构基础。首先，转录调控子的 DNA 结合活性被小分子对其的结合所调节；其次，我们将看到蛋白质-蛋白质相互作用以及蛋白质磷酸化作用是如何介导了一个参与细胞分裂的酶的变构调控。

小分子受动器：由异乳糖调控的 Lac 阻遏蛋白 *E. coli* 的 *lacI* 基因编码了乳糖阻遏蛋白（Lac）。这种蛋白质（我们将在第 16 章具体讨论）由变构作用调控——实际上，它是最早被定义的变构调控 DNA 结合蛋白的实例之一。该蛋白质参与基因调控，当它结合到 DNA 上时，可以阻止基因的转录，该基因的转录是细胞利用乳糖作为碳源所必需的。但是，当环境中存在乳糖时，这种糖的一种特殊形式（β-1-6-allolactose）诱导了乳糖基因的表达。这种异乳糖诱导物通过直接与 Lac 阻遏蛋白结合并使后者与 DNA 的结合不稳定来行使其功能。

结构分析表明，与诱导物结合后 Lac 阻遏蛋白改变了外形［这些结构研究使用了人工诱导物分子异丙基-β-D-硫代半乳糖（IPTG）］。随后这种外形改变解释了为何该蛋白质的 DNA 结合活性减弱了。Lac 阻遏蛋白是一种大蛋白质（155kDa 的四聚物），具有参与 DNA 结合作用、蛋白质多聚作用以及诱导物结合作用的不同结构域。该蛋白质的 N 端区域（1～49 位氨基酸残基）是一个螺旋-转角-螺旋模式，正如我们在 λ 阻遏蛋白

的例子中看到的，该区域特异性的结合于启动子调控区域的 DNA 的大沟。该区域的邻近部分是另一个附加螺旋，被称作铰链螺旋（hinge helix），后者与小沟相互作用。相对的，诱导物结合口袋处于大的核心结构域的中心部分（由第 62～333 位氨基酸残基构成）。

将 LacI 与 DNA 结合时的结构和不与 DNA 结合（与诱导物结合）时的结构相比较，我们可以看出为何这两种状态是如此严格的相互排斥。与诱导物的结合引起了大的核心结构域中 N 端部分排列的扭曲。随之，这种构型改变破坏了铰链螺旋的结构，从而削弱了与 DNA 的结合作用；邻近的螺旋-转角-螺旋结构域的结构被赋予了更多的柔韧性，这一改变会降低蛋白质对其特异性的 DNA 结合位点的亲和性（图 5-26）。

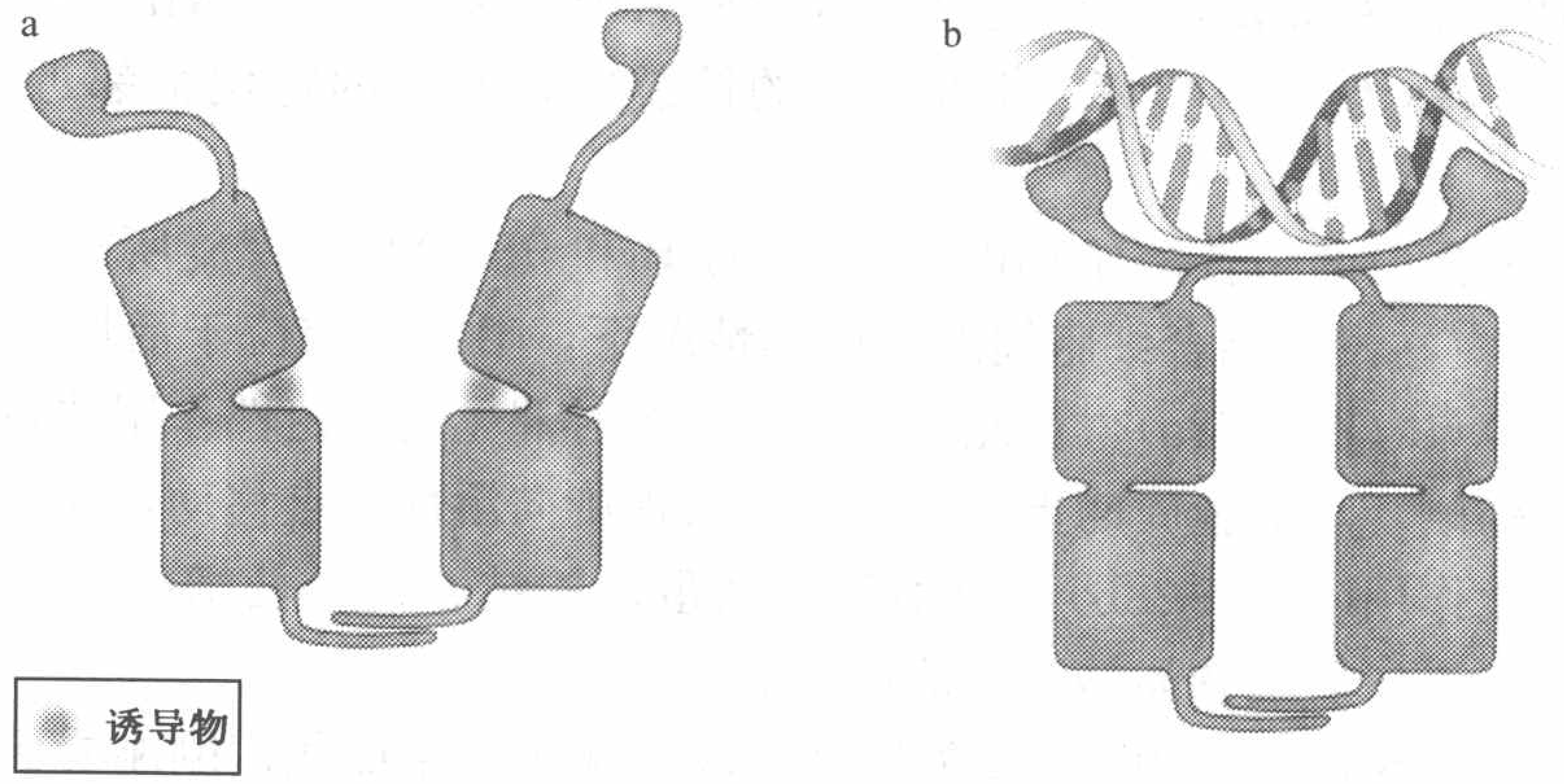


图 5-26　Lac 阻遏蛋白的变构改变。图中每一部分展示的都是一个 Lac 阻遏蛋白的二聚体。a. 上图展示的是诱导物-Lac 阻遏蛋白复合物的二聚体。与诱导物的结合导致结构的改变，从而降低了阻遏蛋白对操作子的亲和性；b. 下图展示的是缺少诱导物的二聚体。在这种情况下，铰链螺旋形成，N 端结构域与操作子序列结合。（来源：Adapted，with permission，from Lewis M. et al. 1996. Science 271：1247-1254，Fig 12. © AAAS.）

天冬氨酸氨甲酰基转移酶由其配体——CTP 介导的变构修饰提供了另一个小分子受动器的实例（图 5-27）。在该例中，配体诱发了蛋白质四级结构的明显改变。

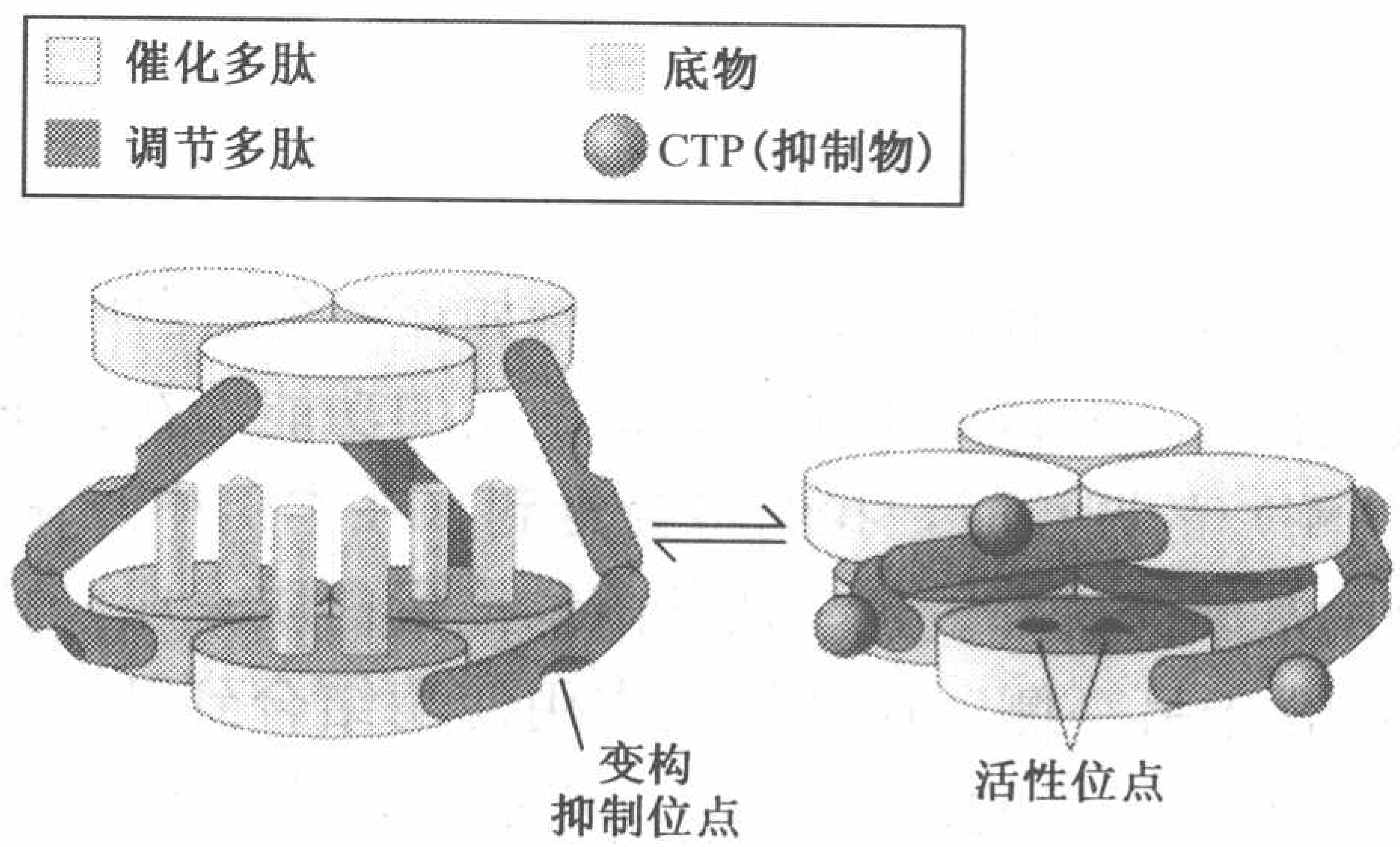


图 5-27　由 CTP 介导的天冬氨酸氨甲酰基转移酶(ATCase)的变构修饰。

蛋白质受动器：由细胞周期蛋白介导的 Cdk 的活化　我们现在转而讨论另一种变构调控，它是通过酶与其调控蛋白间的相互作用实现的。这种酶（被称为 Cdk2）是依赖于细胞周期蛋白的激酶（Cdk）家族的一员。这类激酶调控细胞周期，它只有与被称作细胞周期蛋白的调控蛋白形成复合物才能具有活性。与第二种蛋白质的结合诱导了它构型的改变，从而改变了 Cdk2 活性位点周围的结构，也激活了它的功能。由邻近的特定苏氨酸残基的磷酸化所诱导的进一步的构型变化进一步激活该酶（如下所述）。

与细胞周期蛋白结合所介导的变构过程的具体结构细节已经被确定了。不与细胞周期蛋白结合的 Cdk2 的结构看上去与其他激酶没有什么不同。Cdk2 结构中的两个部分对其调控尤为重要：一个是被称为 PSTAIRE 螺旋（PSTAIRE helix）的 α 螺旋，另一个是被称为 T 噜噗环（T loop）的铰链区；两者的定位都邻近于激酶的活性位点。

与细胞周期蛋白的结合诱导 Cdk2 的 T 噜噗环和 PSTAIRE 螺旋所在位置发生变构改变（图 5-28）。在没有结合细胞周期蛋白的情况下，噜噗环定位于活性位点的入口而螺旋则远离该位点（图 5-28a）。在此构型中，一个对催化起决定性作用的谷氨酸残基处于活性位点的外侧，与细胞周期蛋白的结合导致螺旋向活性位点移动，使得该谷氨酸残基能够参与催化作用（图 5-28b）。与细胞周期蛋白的结合同样使噜噗环从活性位点的入口离开，使得蛋白质底物能够进入活性位点。

作为受动器的磷酸化：由 CAK 介导的 Cdk 的活化　正如我们刚刚看到的，Cdk 通过与细胞周期蛋白结合而被活化，但是 Cdk 的完全活化还需要该酶由磷酸化介导的进一步变构变化。该磷酸化发生在上面提及的 T 噜噗环上的一个苏氨酸残基。这一修饰导致 Cdk 活性位点的进一步重排。一旦磷酸化，磷酸基团就会与 3 个精氨酸结合，后者分别来自催化裂缝周围不同的区域。这种相互作用将催化裂缝固定于一种适合于高活性的构型。

磷酸化反应由另一种激酶（被称为 CAK）催化，许多激酶由相似的磷酸化过程激活。激活 Cdk 需要两个步骤：与细胞周期蛋白的结合及磷酸化，这两个步骤按顺序先后发生。这是因为与细胞周期蛋白的结合不仅仅增加了该酶的活性，而且使 T 噜噗环易于被 CAK 磷酸化。

并非所有的蛋白质调控都由变构介导

某些蛋白质是通过非变构途径进行调控的。例如，一个蛋白质能够将另一个蛋白质结合到其特定部位或底物上，并以此来调控该蛋白质的功能。在讨论 RNA 聚合酶（将基因转录为 mRNA 的酶）的调控时，我们将会看到（在本例中）调控通常意味着对在任一给定时刻所转录的基因的选择。这一任务由调控蛋白质完成，它们结合于 DNA 的一侧而 RNA 聚合酶结合于另一侧。这种相互作用使酶结合到某一（或众多）基因上，从而为特定的调控子提供了正确的结合位点。这是一个蛋白质对 DNA 协同结合（co-operative binding）的实例（第 16 章）。

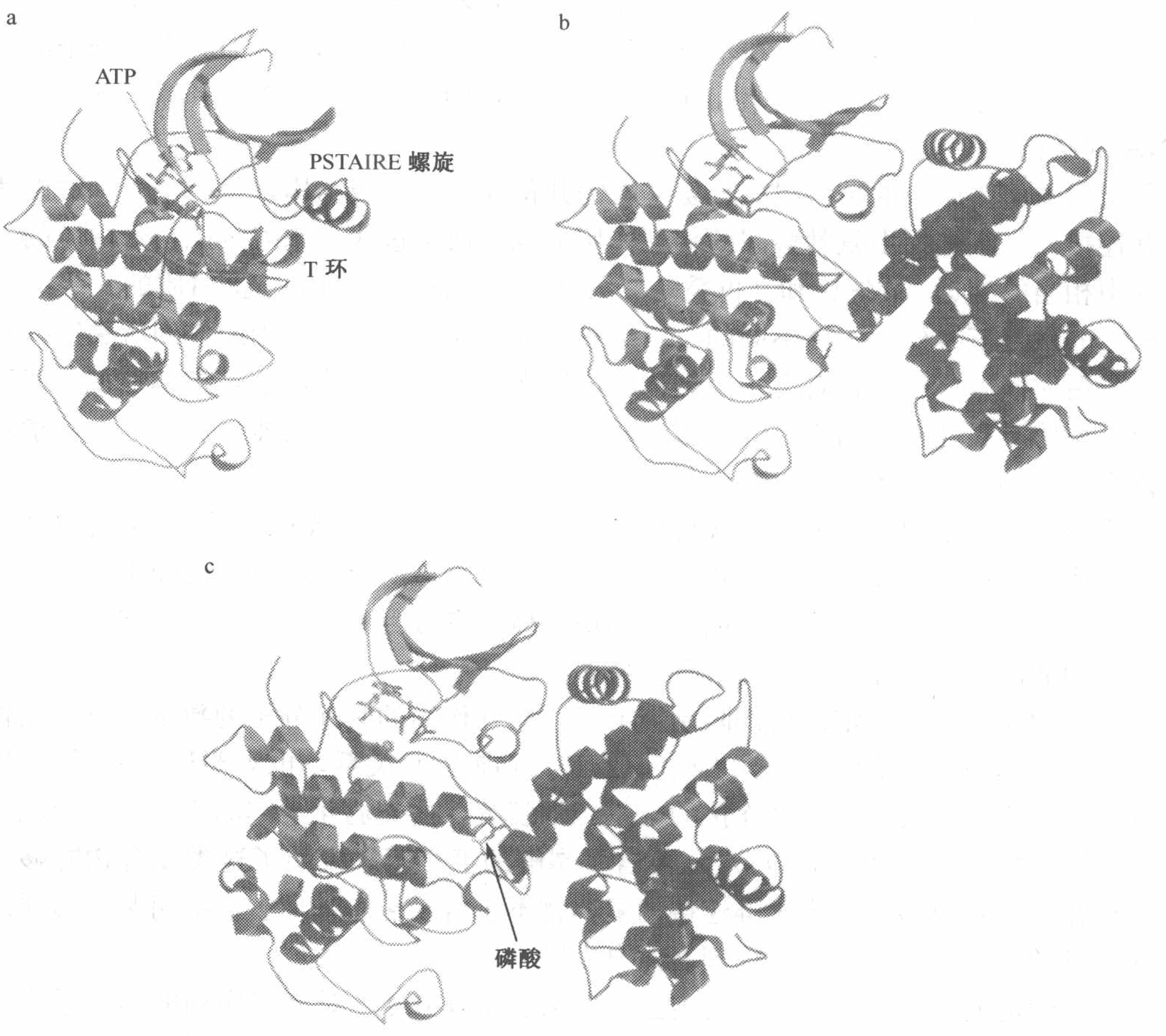


图 5-28 由细胞周期蛋白诱导的 Cdk 的构型变化。 a. 以青色表示的激酶单体结构是无活性的。PSTAIRE 螺旋所在的位置具有一个处于催化中心外侧的氨基酸残基，后者是 ATP 的结合位点，而 T 环阻碍了蛋白质底物（未标出）的进入；b. 该结构展示由于细胞周期蛋白（以紫色表示）的结合，螺旋改变了位置，T 环也从催化中心的入口移开，该复合物具有部分活性；c. 通过 T 环（以红色表示）的磷酸化，Cdk-细胞周期蛋白复合物具有了完全活性。(Schulze-Gahmen U.，DeBondt H.，L.，and Kim S. H.，1996. J. Med Chem 39:4540；Jeffrey P. D.，Russo A. A.，Polyak K.，Gibbs E.，Hurwitz J.，Massague J.，and Pavletich N. P. 1995. Nature 376:313. Russo A. A.，Jeffrey P. D.，and Paveltich N. P.，1996. Nat Scruct Biol 3:696.）镜像图由 MolScript、BobScript 和 Raster 3D 软件制作。

小结

DNA、RNA 和蛋白质都是高分子物质，各自由一套以共价键连接的特定的单体所构成。例如，DNA 由核苷酸链构成，蛋白质由氨基酸链构成。这些高分子的三维外形都是由单体间的多种弱的，或者说次要的相互作用所决定。所以，在 DNA 中，氢键以及核苷酸碱基间的堆积作用造成了该分子的双螺旋特性。同样地，一个给定蛋白质稳定

的三维结构需要多肽链中氨基酸（不相邻的）间的多种弱相互作用。这类弱化学键的本质我们已经在第 3 章讨论过；在本章，我们讨论这类弱相互作用是如何决定分子的外形以及分子间和分子内的相互作用，重点是蛋白质（我们将在第 6 章具体讨论 DNA 和 RNA 的结构）。

蛋白质由多层次的结构组成。氨基酸的共价连接是一级结构，每一个氨基酸以肽键相互连接。二级结构由氨基酸间的相互作用形成，通常这些氨基酸彼此在蛋白质的一级结构中相当的接近。α 螺旋和 β 折叠是二级结构元件的两个例子。蛋白质的三级结构是多肽链的最终三维外形，由不同的二级结构元件以能量有利为原则的排列所决定。对于许多蛋白质来说，还存在着另一个结构组成层次——四级结构，这是指单独多肽链复合组成二聚体或高度有序结构。许多蛋白质以多聚体形式行使其功能。例如，血红素是四聚体而许多 DNA 结合蛋白是二聚体。

许多天然的蛋白质具有一些各自独立的折叠区域（结构域），它们由连续的氨基酸序列形成，在单独存在的情况下相当稳定。这类结构域的组合造成了已知蛋白质的丰富多样性。实际存在的具有唯一特征的结构域大概仅有数百种，每一种结构域经常与一种特定的功能活性联系在一起，如结合 DNA。

每一种大分子的特定外形都限制了能与之相互作用的其他分子的数量。分子间较强的次要相互作用既需要两个结合表面间的互补（锁-钥）关系，也需要许多原子的参与。尽管仅由一或两个次要化学键连接的分子经常裂解，这类弱化学键的集合却能形成一个相当稳定的复合物。双螺旋 DNA 从不自发裂解的事实正好说明了这类复合物能够如何稳定存在。尽管由多个弱化学键连接的复合物不会自发裂解，但它们的组装能够通过正确化学键的逐步形成（自组装原理）而自发发生。

弱化学键的形成同样参与了特定蛋白质沿着 DNA 分子对特定序列的结合过程。这类弱化学键通常是 DNA 碱基基团与蛋白质上的适当供体或受体基团形成的氢键。大多数调控蛋白运用 α 螺旋识别特定的 DNA 序列。这种“识别螺旋”插入 DNA 的大沟，螺旋中的氨基酸序列特异性的结合碱基的边缘。这种结合被特异性相互作用的结合能量所稳定。DNA 结合蛋白同样具有允许其与 DNA 骨架非特异性结合的区域。这种非特异性的骨架相互作用使蛋白质沿着 DNA 线性迁移，使其能够更快的抵达其特异性的靶序列。极少数蛋白质运用 β 折叠（而非 α 螺旋）识别特定的 DNA 序列，与小沟相互作用，但这种情况并不多见。

蛋白质行使多种功能，诸如催化或结合 DNA。这些活性一般由以下几种机制调控，包括小分子配体或其他蛋白质对目标蛋白的结合，以及蛋白质中残基的酶促修饰。这些配体或修饰，经常通过变构调控蛋白质的功能。也就是说，配体结合（或靶点修饰）于蛋白质上的某个位点，该位点与蛋白质行使其主要功能的区域（如酶的活性位点、DNA 结合结构域等）相分离。这种结合或修饰造成了蛋白质外形的改变，从而增强或减弱了活性位点或 DNA 结合结构域的活性，从本质上开启或关闭了蛋白质的活性。

另一些例子说明，蛋白质可以被第二个蛋白质的修饰或结合所调节，从而不需要变构的参与。例如，修饰可以在蛋白质上形成一个能够被第二个蛋白质识别的位点。这类

蛋白质-蛋白质相互作用能够将蛋白质结合到其特定部位或底物上，并以此来调控其活性。

（时亮 译 吴东颖 刘国振 刘韧 校）

参考文献

书

Branden C. and Tooze J. 1999. *Introduction to protein structure*, 2nd ed. Garland Publishing, New York.

Pauling L. 1960. *The nature of the chemical bond*, 3rd ed. Cornell University Press, Ithaca, New York.

Petsko G.A. and Ringe D. 2003. *Protein structure and function (primers in biology)*. New Science Press, Waltham, Massachusetts.

氢键的排列方式决定蛋白质的特定构型

Chothia C. 1984. Principles that determine the structures of proteins. *Annu. Rev. Biochem.* **53:** 537–572.

大多数蛋白质由两个或三个结构域装配而成

Rose G.E. 1979. Hiererarchical organization of domains in globular proteins. *J. Mol. Biol.* **134:** 447–470.

Steitz T.A., Weber I.T., and Matthew J.B. 1982. Catabolite gene activator protein: Structure, homology, with other proteins, and cyclic AMP and DNA binding. *Cold Spring Harbor Symp. Quant. Biol.* **47:** 419–426.

弱化学键决定蛋白质在DNA和RNA分子上的正确位置

Auweter S.D., Oberstrass F.C., and Allain F.H.-T. 2006. Sequence-specific binding of single-stranded RNA: Is there a code for recognition? *Nucleic Acids Res.* **34:** 4943–4959.

De Guzman R.N., Turner R.B., and Summers M.F. 1999. Protein–RNA recognition. *Biopolymers* **48:** 181–195.

Lunde B.M., Moore C., and Varani G. 2007. RNA-binding proteins: Modular design for efficient function. *Nat. Rev. Mol. Cell Biol.* **8:** 479–490.

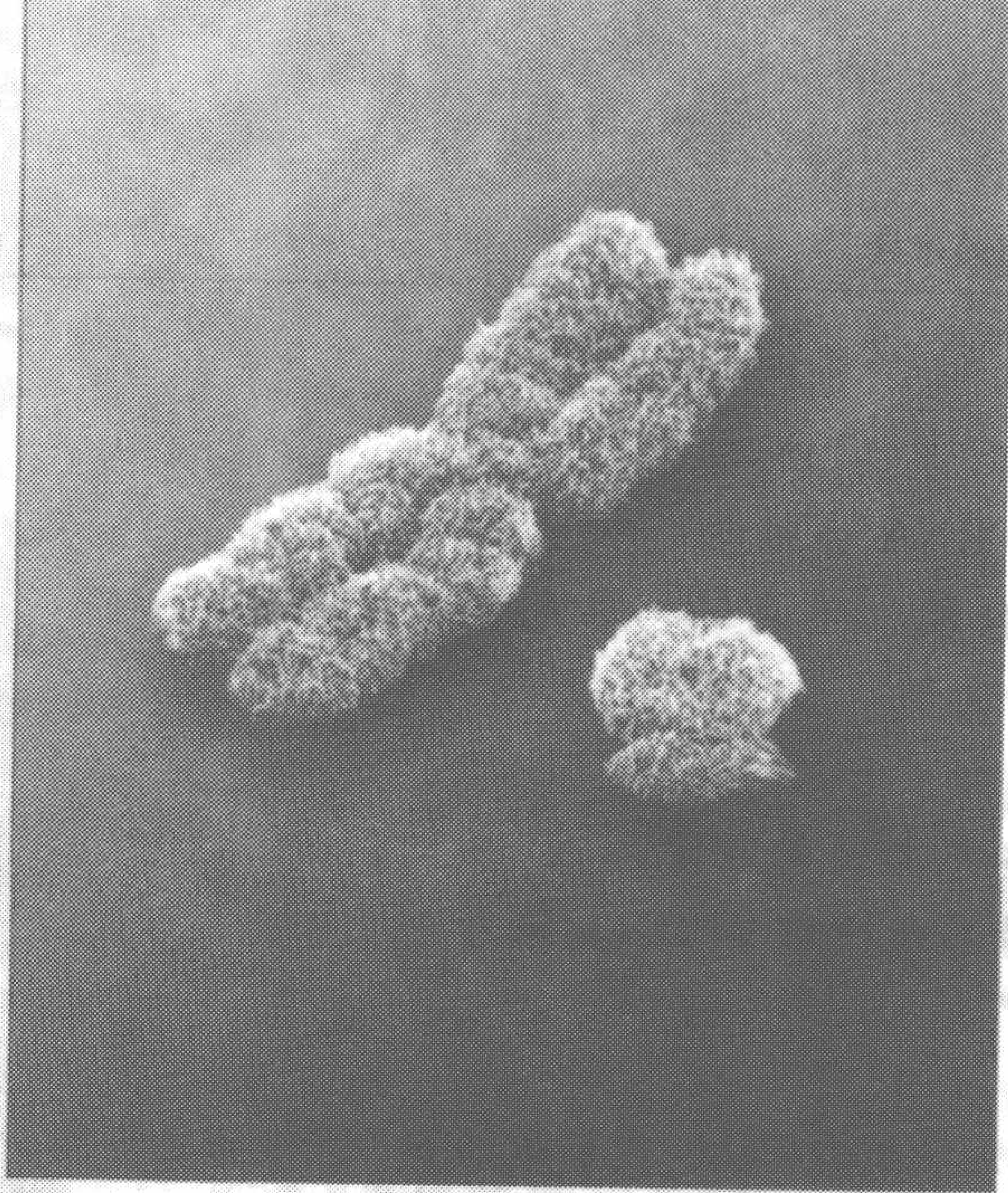
Sperling R. and Wachtel E.J. 1981. The histones. *Adv. Prot. Chem.* **34:** 1–52.

变构：通过改变蛋白质的外形调节其功能

Bell C.E. and Lewis M. 2001. The Lac repressor: A second generation of structural and functional studies. *Curr. Opin. Struct. Biol.* **11:** 19–25.

Pace H.C., Kercher M.A., Lu P., Markiewicz P., Miller J.H., Chang G., and Lewis M. 1997. Lac repressor genetic map in real space. *Trends Biochem. Sci.* **22:** 334–338.

Pavletich N.P. 1999. Mechanisms of cyclin-dependent kinase regulation: Structures of cdks, their cyclin activators, and cip and INK4 inhibitors. *J. Mol. Biol.* **287:** 821–828.



第2篇 基因组的维持

本篇概要

- 第 6 章 DNA 和 RNA 的结构
- 第 7 章 基因组结构、染色体、染色质和核小体
- 第 8 章 DNA 的复制
- 第 9 章 DNA 的突变和修复
- 第 10 章 分子水平上的同源重组
- 第 11 章 位点特异性重组和 DNA 转座

第 2 篇将介绍 DNA 的结构和从一代细胞到下一代细胞的复制、遗传和变异过程。在第 6～11 章中，我们将仔细分析 DNA 和与它关系密切的 RNA，着重探讨以下问题：

- DNA 和 RNA 分子的结构是如何影响其功能的?
- 组成真核生物染色体的、比细胞大小要长很多的 DNA 分子怎样被包装在细胞核里?
- DNA 分子在细胞周期中如何进行精确而完整的复制，如何达到复制的高保真性?
- DNA 分子如何避免来自于自身和环境的损伤，受损后怎样进行修复?
- 重组和转座过程中染色体间 DNA 片段的交换如何进行，又有什么样的生物学意义?

在回答这些问题的过程中，我们将注意到 DNA 分子既受到保守作用的影响——使其在代代相传的过程中保持不变；同时又受到相反作用的影响，使遗传物质发生重要变化，推动生物的多样性和进化。DNA 在细胞中受不同力的作用，造成一系列不同的结果，包括解链，形成更紧密的拓扑结构，相互缠绕并与蛋白质结合，以及分子骨架断裂和重新连接。这些作用是由复制、保持和改变遗传物质的多种酶和其他分子介导的。

第 6 章详细地从原子水平讲述了 DNA 的结构，包括从碱基和骨架的化学组成到碱基配对作用以及其他维持双链的作用。DNA 通常处于受限的拓扑结构，该章节介绍了这种限制的生物效应，以及改变拓扑结构的酶类。除了 DNA 外，RNA 的结构也在第 6 章做了介绍。尽管 RNA 的化学组成与 DNA 非常相似，但具有自己独特的结构和特性，包括作为催化剂在一些细胞活动过程中起作用。

第 7 章描述组成染色体的巨大 DNA 分子在不同生物体中不同的组织和大小巨大的染色体 DNA 不是裸露的，它通过和一些特殊的蛋白质一起包装形成称为染色质的结构。这一包装过程可以使细胞容纳下很长的 DNA 分子，不仅使染色体的长度变短，而且改变了 DNA 的可接近性和行为。染色质经修饰后，可以使 DNA 变得更易于或难以接近。DNA 的这些变化可以确保它在正确的时间和位置进行复制、重组和转录。第 7 章还介绍了染色质的组蛋白和非组蛋白组分、染色质结构以及调节染色体 DNA 可接近性的各种酶。

从 DNA 的结构可以推测出遗传物质复制的可能机制。第 8 章详细描述了 DNA 的复制机制，以及保证 DNA 双链同时复制所需的多种酶和复杂的分子机器。我们还讨论了 DNA 复制过程是如何启动的，细胞又是如何精密调控 DNA 的复制，来维持合适数目的染色体。

DNA 的复制并不是万无一失的，每一轮复制出错后若不纠正，都会导致子代 DNA 分子发生突变。而且，DNA 分子是脆弱的，会自发发生损伤，或受到化学物质或辐射的损伤。如果想要避免遗传物质的恶性突变的快速积累，就需要检测并修复这些损伤。第 9 章介绍了 DNA 损伤的检测和修复机制。从细菌到人类，各种生物体都依赖于相似的高度保守的 DNA 损伤的检测和修复机制来保证 DNA 的完整性，DNA 损伤的检测和修复系统的失常将导致可怕的后果，如癌症。

第 2 篇的最后两章介绍了 DNA 代谢的另一个方面。与确保遗传物质的变化降低到最小限度的保守的复制和修复过程相比，这些过程会引起 DNA 序列的重新排列。第 10 章介绍了同源重组——同源染色体通过断裂、重新连接以互换相应的 DNA 片段的过程。作为 DNA 损伤的主要修复机制，同源重组丰富了遗传多样性，并使缺失或受损的序列得到替换。该章还详细介绍了同源重组途径的两种模式，同时也介绍了一系列分子“机器”——它们在 DNA 分子间搜寻同源序列，生成并消除两种同源重组模式中的中间产物。

最后，第 11 章介绍了两种特殊的重组——位点特异性重组和转座，这些过程会导致许多生物体包括人类基因组中的某些序列的大量积聚。我们还将在本章讨论这些遗传物质交换的分子机制及其产生的生物学影响。

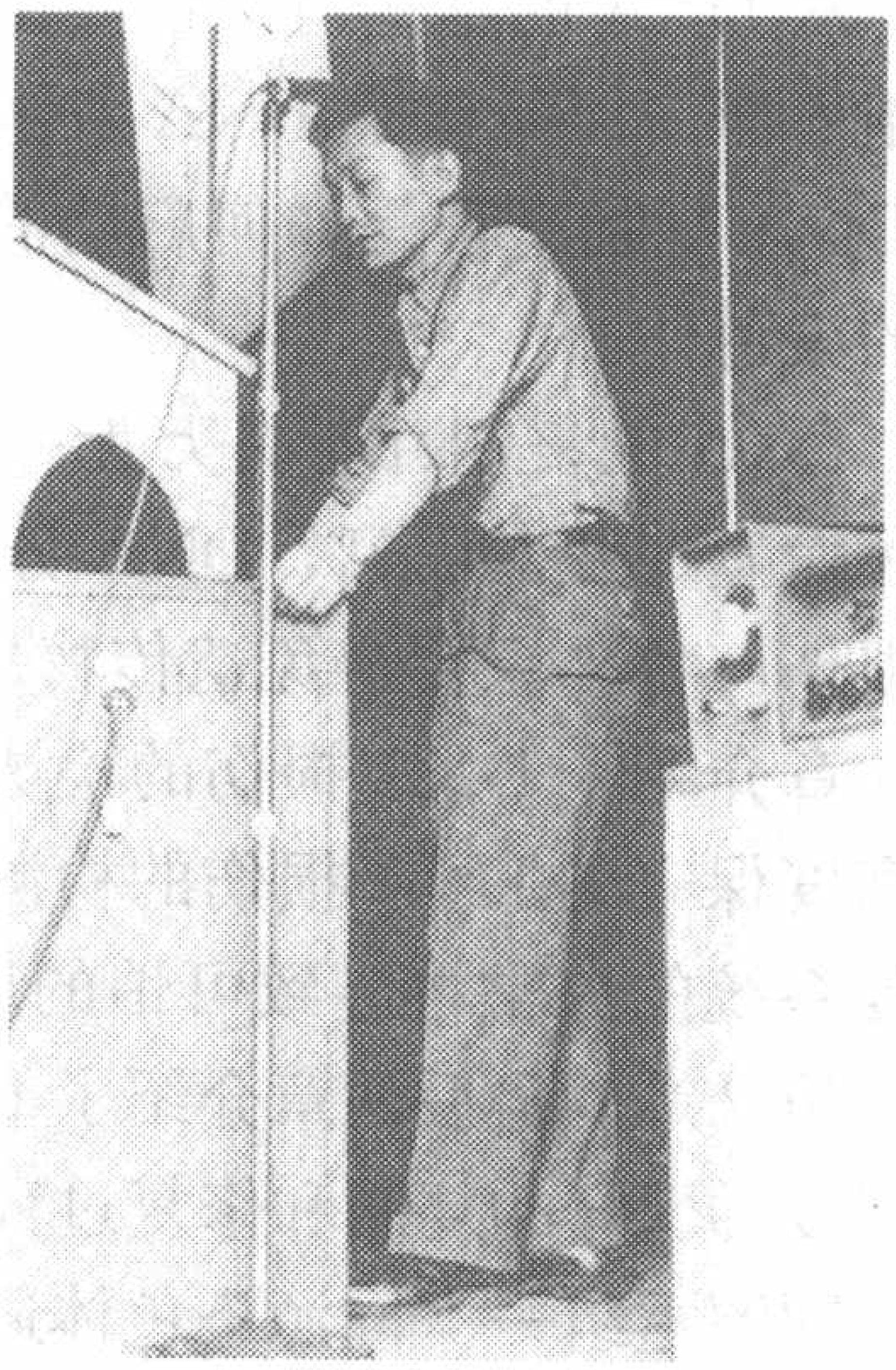
冷泉港实验室档案库中的照片



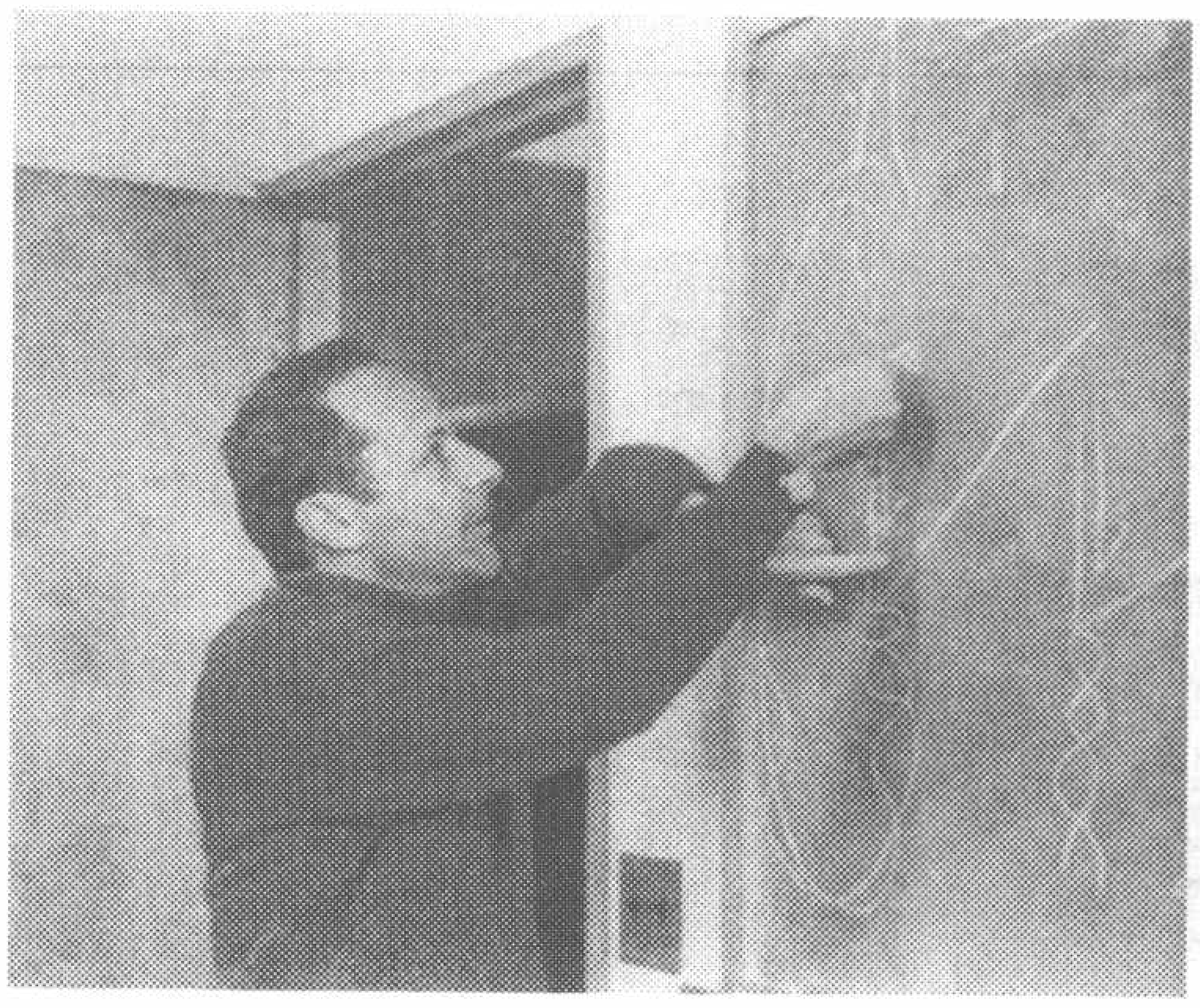
Barbara McClintock 和 Robin Holliday，1984 年的 DNA 水平重组学术讨论会。McClintock 提出转座子的存在以解释她在 20 世纪 40 年代开展的玉米遗传学的研究结果（第 11 章）。这一成果直到 30 年后的 1983 年才被肯定并荣获诺贝尔生理学或医学奖。Holliday 则提出了以他名字命名的同源重组的基本模型（第 10 章）。



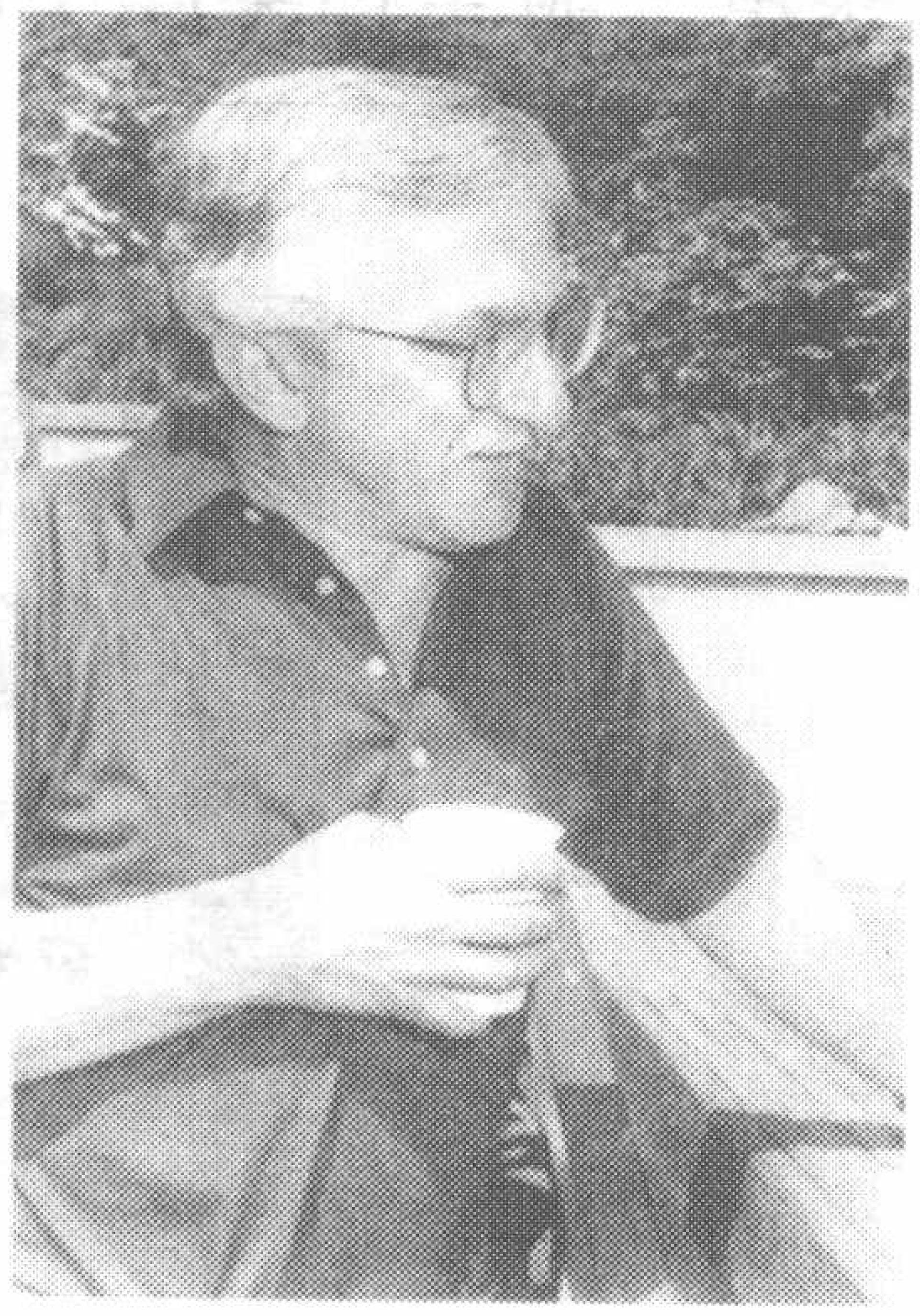
Arthur Kornberg，1978 年 DNA 复制和重组研讨会。 Kornberg 对在研究 DNA 复制（第 8 章）领域有突出的贡献。他纯化出第一个可以合成 DNA 的酶，这是一种来源于大肠杆菌的 DNA 聚合酶。他用实验证明新的 DNA 合成需要 DNA 模板，从而证实了沃森和克里克提出的 DNA 复制模型。由于上述的工作，Kornberg 与 Severo Ochoa 分享了 1959 年的诺贝尔医学或生理学奖。



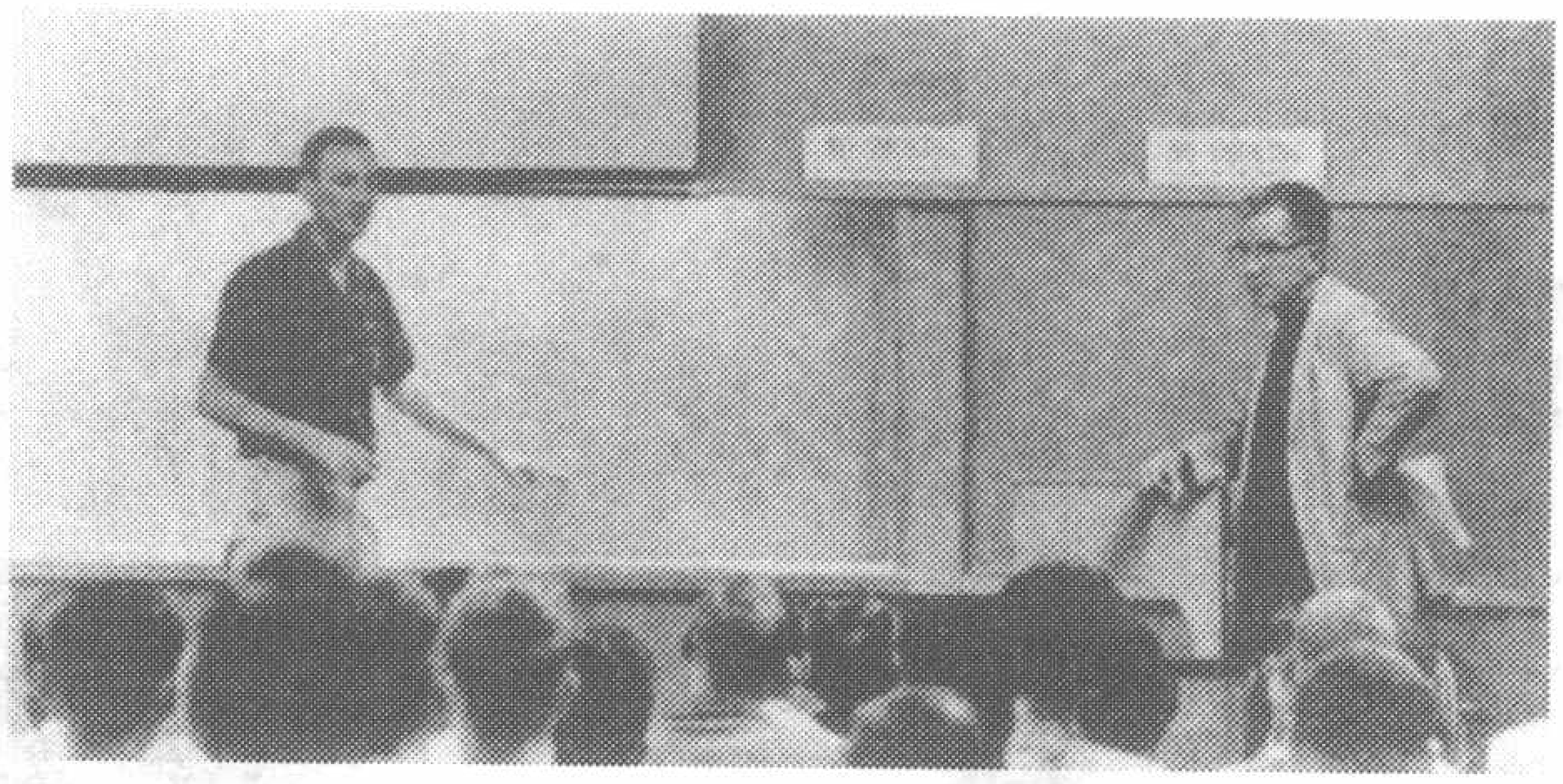
Reiji Okazaki，1968 年微生物 DNA 复制研讨会。 Okazaki 此时刚刚揭示了在 DNA 复制时其中一条新生链是如何先以小片段的方式开始合成，然后再连接起来的。“Okazaki 片段”解释了只能以一个方向合成的酶为什么却能同时生成两条极性相反的链（第 8 章）。



Matthew Meselson，1968 年微生物 DNA 复制研讨会。 Meselson 和 Frank Stahl 一起用实验证明了 DNA 的半保留复制——“最亮”（见下面的照片和第 2 章）Meselsen 对数个领域都作出了重要贡献，包括纯化第一个限制性内切酶，在这一张照片拍摄时出版。他为防止化学与生物武器的工作广为人知。



保罗·默德里希（Paul Modrich），1993 年 DNA 和染色体研讨会。 Modrich 是 DNA 修复领域的先驱，他揭示了错配修复的大部分机制（第 9 章）。



Franklin Stahl 和 Max Delbrück，1958 年的遗传物质交换：机制与结果研讨会。Stahl 与 Meselson（见上照片）一起。证明了 DNA 的半保守复制。这一实验曾被称为"最漂亮的生物学实验（第 2 章）。Stahl 随后为我们理解同源重组作出了很大的贡献（第 10 章）。Delbruck 是"噬菌体小组"（一班在冷泉港实验高度夏的科学家，将噬菌体发展成生物学的第一个模式系统的最有影响的创始人之一（第 22 章）。



Carol Greider、Titia de Lange 和 Elizabeth Blackbum，2001 年端粒会议。Blackbum 发现了在染色体末端的端粒重复的序列特征。后来，作为 Blackbum 实验室的研究生，格雷德发现了端粒酶。这个酶可以维持端粒（第 8 章）。在她们中间的是 de Lange，她的工作主要集中于细胞中结合并保护端粒的蛋白质。



Jack Szostak（右），1981 酵母会议。 Szostak 与 Blackbum（上）莱斯泰克，发现了在四膜虫（Tetrahymena）中端粒序列。当端粒序列结合到酵母的线性 DNA 分子的末端，可以保护 DNA。他们认为是一种酶——后来由 Greide（上）发现的——形成的端粒（第 8 章）。Szostalk 正在与一个人说话；背景是 Jim Hicks，当时他和 Jeff Strathem 和 Amar Klar 正在冷泉港实验室工作，研究酵母接合型转变的机制（第 11 和 17 章）。

（李京湘　译　胡松年　刘韧　杨焕明　校）

第6章 DNA 和 RNA 的结构

本章概要

•

DNA 的结构

•

DNA 拓扑学

•

RNA 的结构

人们在发现 DNA 是携带了染色体所有遗传信息的主要遗传分子后，立即将注意力集中在 DNA 的结构上。当时，人们希望通过了解 DNA 结构，能够揭示染色体分开并形成两个相同的拷贝时，DNA 携带的遗传信息是如何进行复制的。从 20 世纪 40 年代末到 50 年代初，几个美国和欧洲的研究小组既协作又竞争地对此进行了研究，以期了解 DNA 的原子间是如何通过共价键连接在一起，而由此形成的 DNA 分子是如何在三维空间中排列的。不足为奇，刚开始人们担心 DNA 的结构非常复杂甚至可能很古怪，不同基因中的 DNA 结构是完全不同的。因此，当 DNA 的双螺旋结构被发现时，人们大大松了一口气。DNA 的这一基本结构告诉我们，所有基因的三维结构都大致相同，区别只在于 DNA 互补双链中 4 种核苷酸的数量和排列顺序的不同。

双螺旋结构的发现距今已有 50 多年，关于遗传物质的这一阐述仍然未受到质疑，不需要做改动以适应新的发现。然而，人们也逐渐认识到 DNA 的结构并不完全和刚开始想的完全一致。例如，一些小病毒的染色体 DNA 是单链而不是双链分子。而且，受局部 DNA 序列的影响，不同的碱基对的精确方向有着细微的差别。一些序列甚至能使 DNA 卷曲成左手双螺旋，与最初阐述的 DNA 基本结构——右手双螺旋正好相反。虽然有些 DNA 分子是线性的，但也有一些是环状的。DNA 双螺旋经常围绕 DNA 结合蛋白质进一步盘绕卷曲成超螺旋结构，这样造成了更大的复杂性。

同样，我们现在认识到 RNA 虽然乍看起来与 DNA 非常相似，但是它有自己独特的结构特征。虽然 RNA 主要以单链分子的形式存在，但是通过链内碱基配对，也呈现出很多双螺旋的特征，并能折叠成大量不同的三级结构。这些 RNA 结构有诸多令人惊奇之处，例如，非经典碱基对，碱基和骨架的相互作用和结点状构型。最值得一提的是，一些 RNA 分子是酶类，催化从核酸到蛋白质的信息传递中的核心反应，具有重要的进化意义。

显然，DNA 和 RNA 的结构比开始所设想的更加丰富多样，更加复杂。实际上，不存在一种适用于 DNA 和 RNA 的通用结构。本章将要讨论，多核苷酸链由于其独特的物理、化学、拓扑学特性，在一般结构的基础上能发生很多变化。

DNA 的结构

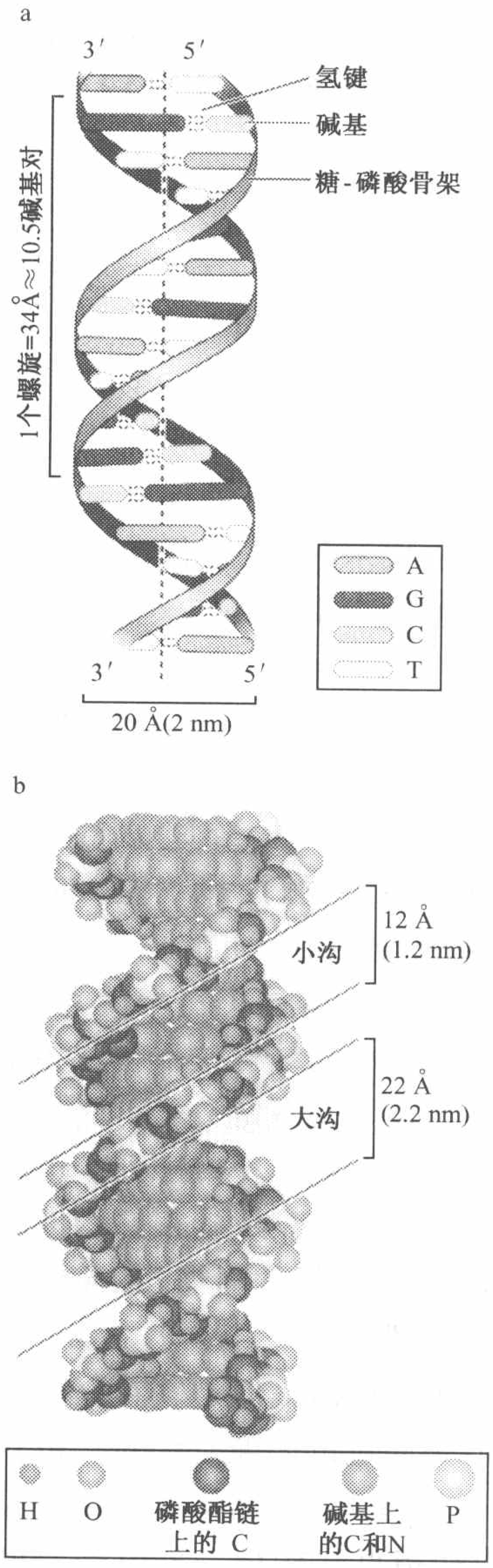


图 6-1 DNA 的双螺旋结构。 a. DNA 双螺旋的结构示意图。螺旋的每一圈（34Å 或 3.4nm）大约有 10.5 个碱基对。b. 双螺旋的空间结构实体模型。糖和磷酸基团构成了每条链的骨架，以黄、灰和红圈表示，（本图）显示了 DNA 分子的螺旋情况，碱基朝向内部，但可以通过大、小沟而被接近。

DNA 由多核苷酸链构成

通常 DNA 最重要的结构特征是两条多核苷酸链（polynucleotide chain）以双螺旋的形式互相缠绕（图 6-1；参见结构指南 6-1）。图 6-1a 为双螺旋结构的示意图，由于 DNA 双链互补，如果将它倒置 180°（如将这本书上下颠倒），双螺旋看起来仍跟原来是一样的，这是由双链的互补性造成的。图 6-1b 是双螺旋的分子结构空间实体模型，体现了 DNA 分子的组成部分及在螺旋结构中的相对位置。双螺旋两条单链的骨架是由糖和磷酸基团交替构成的；碱基朝向骨架的内侧，通过大沟和小沟可以接近碱基。

让我们首先来看 DNA 最基本的组成成分——核苷酸的特点。一个磷酸基团和戊糖相连形成2′-脱氧核糖（2′-deoxyribose），一个碱基附着于2′-脱氧核糖即构成了核苷酸。磷酸基团和戊糖的结构如图 6-2 所示。之所以叫做 2′-脱氧核糖是因为戊糖的第二位碳原子没有羟基（而是两个氢）。请注意核糖上

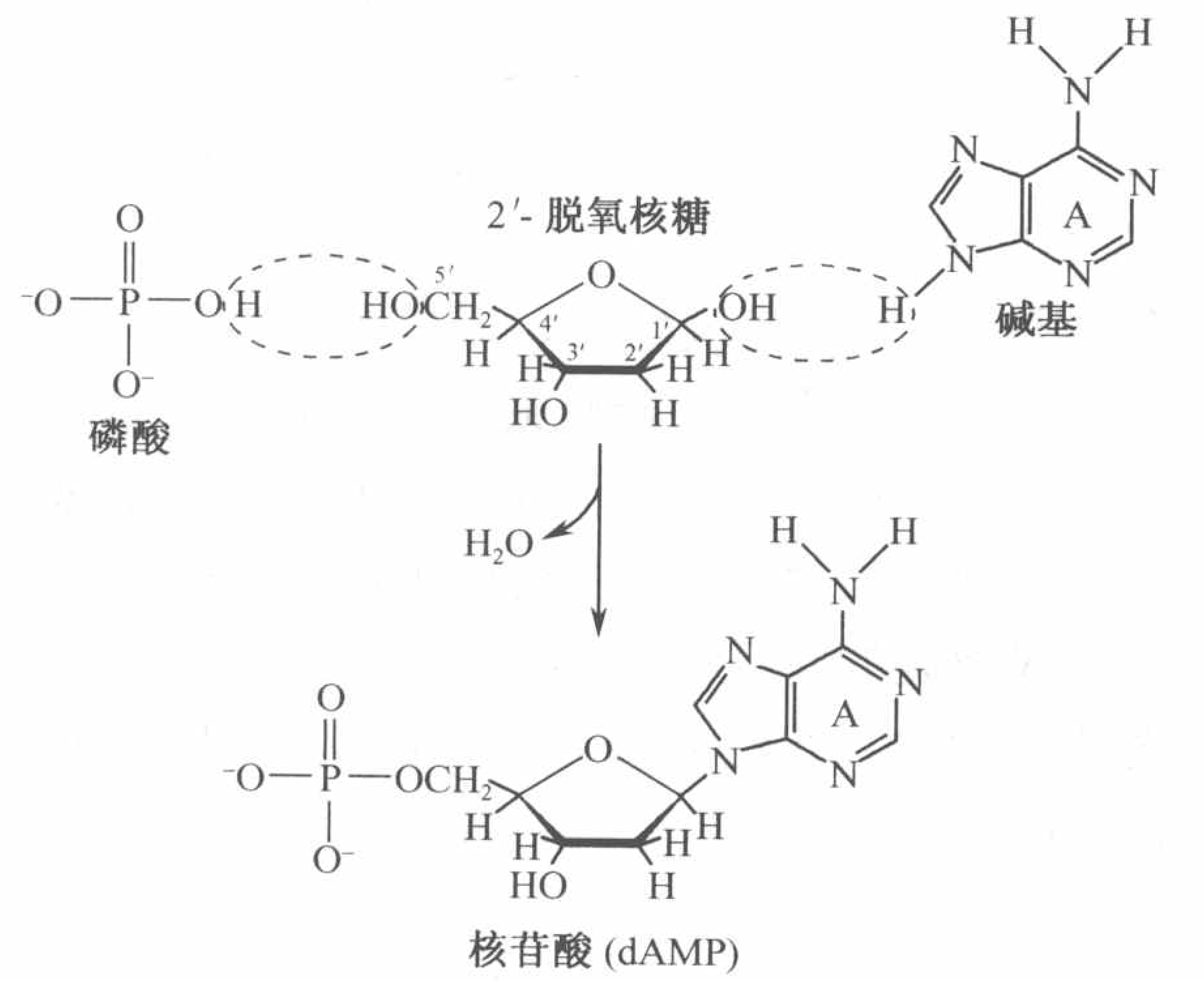
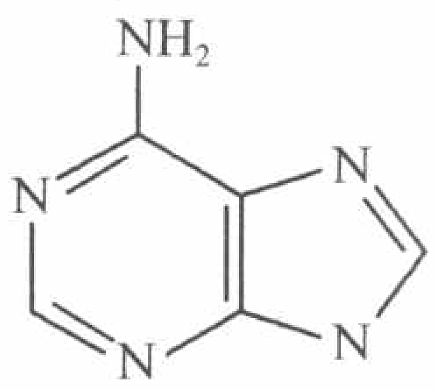
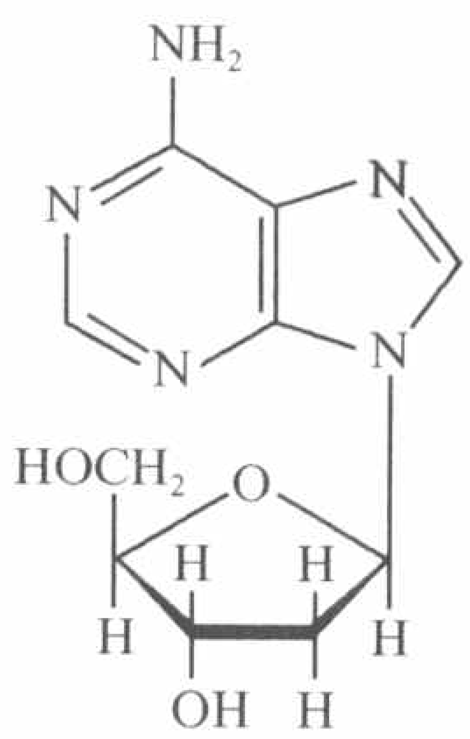
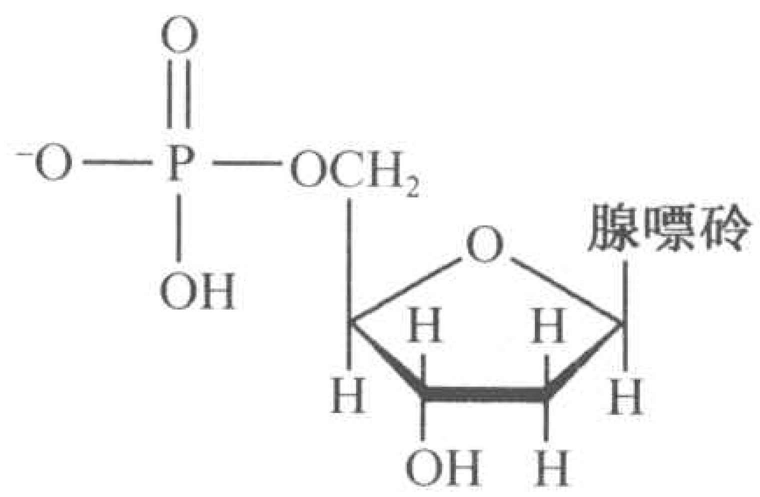


图 6-2 脱水形成核苷酸。 在 2′-脱氧核糖的 C 原子的数目以红色表示。

原子的位置在右上角都标以“′”以区别于碱基上原子的位置（详见下述讨论）。

我们可以想像，碱基通过与戊糖 1′位碳原子上的羟基结合脱去一分子水形成糖苷键（图 6-2）因而与 2′-脱氧核糖结合。戊糖和碱基相结合形成核苷（nucleoside）。同样地，我们可以想像磷酸基与 2′-脱氧核糖 5′位碳原子上的羟基结合脱水形成 5′-磷酸单酯。核苷加一个（或一个以上）磷酸形成核苷酸（nucleotide）。因此，碱基通过糖苷键与糖相连，而糖又通过磷酸酯键与磷酸相连，这样就形成了核苷酸（表 6-1）。

表 6-1　腺嘌呤与相关复合物

	碱基 腺嘌呤	核苷 2′-脱氧腺苷	核苷酸 2′-脱氧腺苷-5′磷酸
结构			
相对分子质量	135.1	251.2	331.2

核苷酸依次以磷酸二酯键相连形成多核苷酸链，即一个核苷酸 2′-脱氧核糖上的 3′-羟基与另一个核苷酸上连接于 5′-羟基的磷酸基相连（图 6-3）。在磷酸二酯键（phosphodiester linkage）中，两个核苷酸间的磷酰基的一个糖通过 3′-羟基酯化，另一个糖通过 5′-羟基酯化。多核苷酸链以磷酸二酯键为基础构成了规则的不断重复的糖-磷酸骨架，这是 DNA 结构的一个特点。与此相反，多核苷酸链中碱基的排列顺序却是无规律的，碱基排列顺序的不规则性加上其长度是 DNA 储存大量信息的基础。

磷酸二酯键赋予了 DNA 链的固有极性，这种极性由核苷酸的不对称性和它们结合的方式所决定。DNA 链的一端是游离的 5′-磷酸基或 5′-羟基，另一端是游离的3′-磷酸基或 3′-羟基。习惯上，DNA 序列由 5′端（在左边）向 3′端书写，一般 5′端是磷酸基而 3′端是羟基。

每个碱基都有它首选的异构体

构成 DNA 的碱基可以分为两类，嘌呤（purine）和嘧啶（pyrimidine）。嘌呤为腺嘌呤（adenine）和鸟嘌呤（guanine）；嘧啶为胞嘧啶（cytosine）和胸腺嘧啶（thymine）。图 6-4 显示了嘌呤的双环结构，腺嘌呤和鸟嘌呤具有这种相同的基本结构，但附着的基团不同；同样，图 6-4 也显示出了胞嘧啶和胸腺嘧啶的单环结构的差异。图 6-4 还用数字标注了嘌呤和嘧啶环上的位置，碱基在嘧啶的 N1 位置上通过糖苷键与脱氧核糖相连而在嘌呤的 N9 位置上通过糖苷键与脱氧核糖相连。

每个碱基都存在两个可选的异构体（tautomeric state），它们互相保持平衡。这种

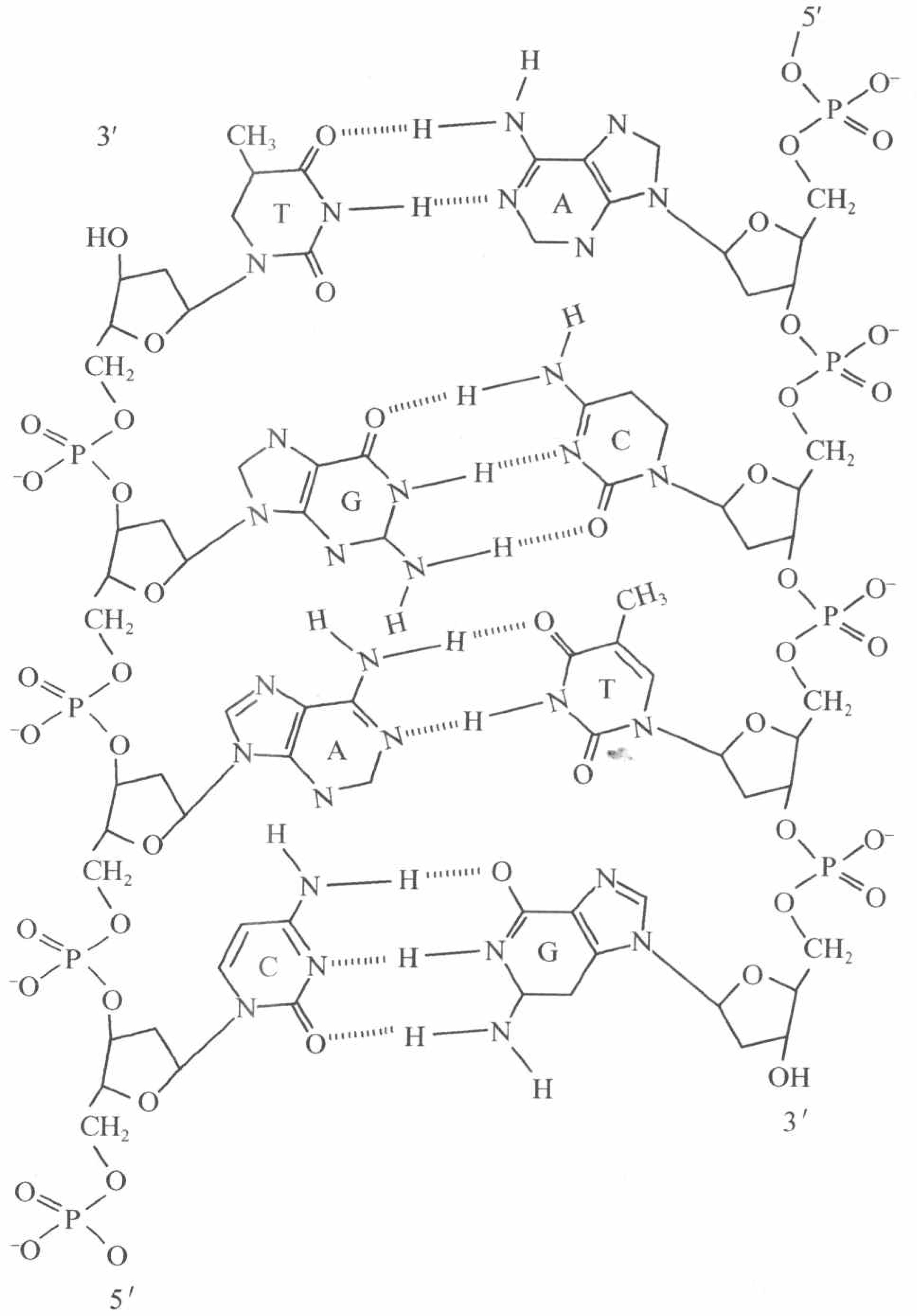
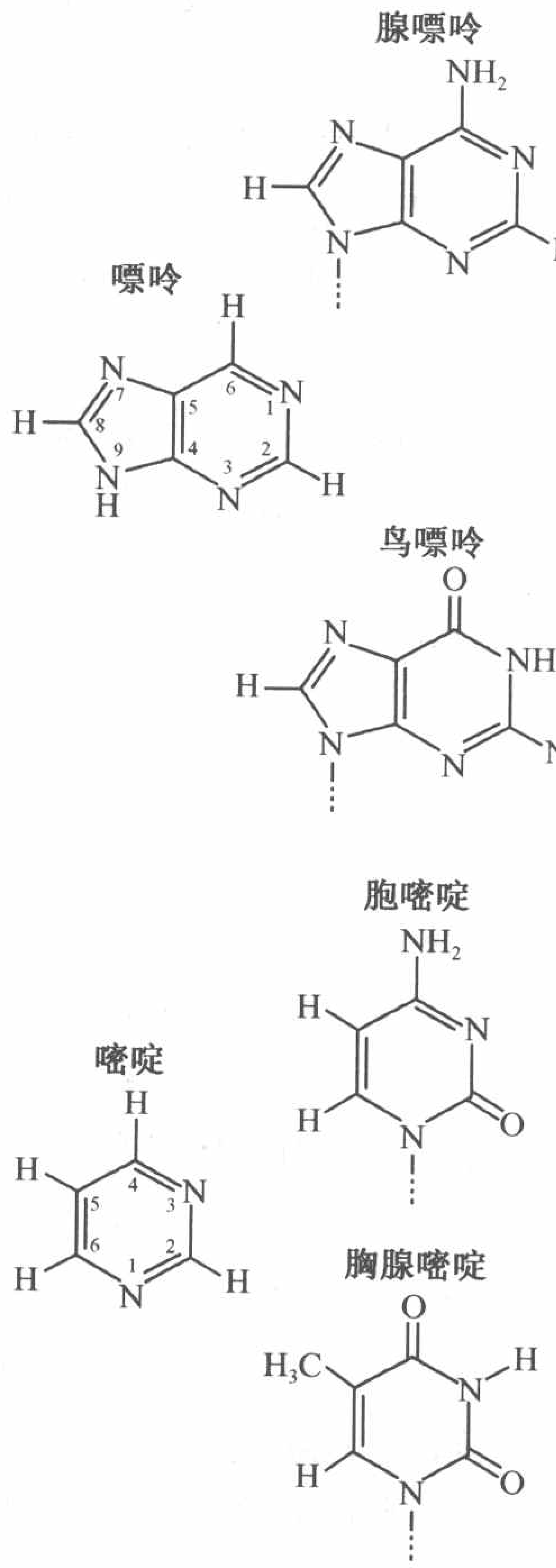


图 6-3 多核苷酸链的详细结构。本图表示嘌呤（蓝色）和嘧啶（黄色）之间的碱基配对作用，以及骨架间的磷酸二酯键。（来源：Dickerson R. E. 1983. Scientific American 249：94，绘图：Irving Geis. 由 Irving Geis Collection/Howard Hughes Medical Institutio 提供，未经许可不得复制）

平衡极大地倾向于图 6-4 所示的常规结构，它们是碱基配对的主要结构。嘌呤和嘧啶环上的氮原子大多以氨基形式存在而很少以亚氨基构型出现；同样，鸟嘌呤和胸腺嘧啶上的氧原子通常以酮式存在而很少以烯醇式构型出现。例如，图 6-5 显示了胞嘧啶的亚氨基异构体，图 6-5b 为鸟嘌呤的烯醇式异构体。像我们看到的，具有形成异构体的能力是导致 DNA 合成时出错的普遍来源。

双螺旋的两条链通过碱基反向平行配对连接在一起

两条多核苷酸链上的碱基靠较弱的、非共价键结合使其互相缠绕形成双螺旋，如图 6-3。一条链上的腺嘌呤总是和另一条上的胸腺嘧啶配对，而鸟嘌呤总是和胞嘧啶配对。两条链有相同的几何螺旋结构，但连接它们的碱基配对使得它们有相反的极性，即一条链上 5′端的碱基和另一条链上 3′端的碱基配对，叫做反向平行，这是腺嘌呤和胸腺嘧啶、鸟嘌呤和胞嘧啶互相配对的立体化学结果。



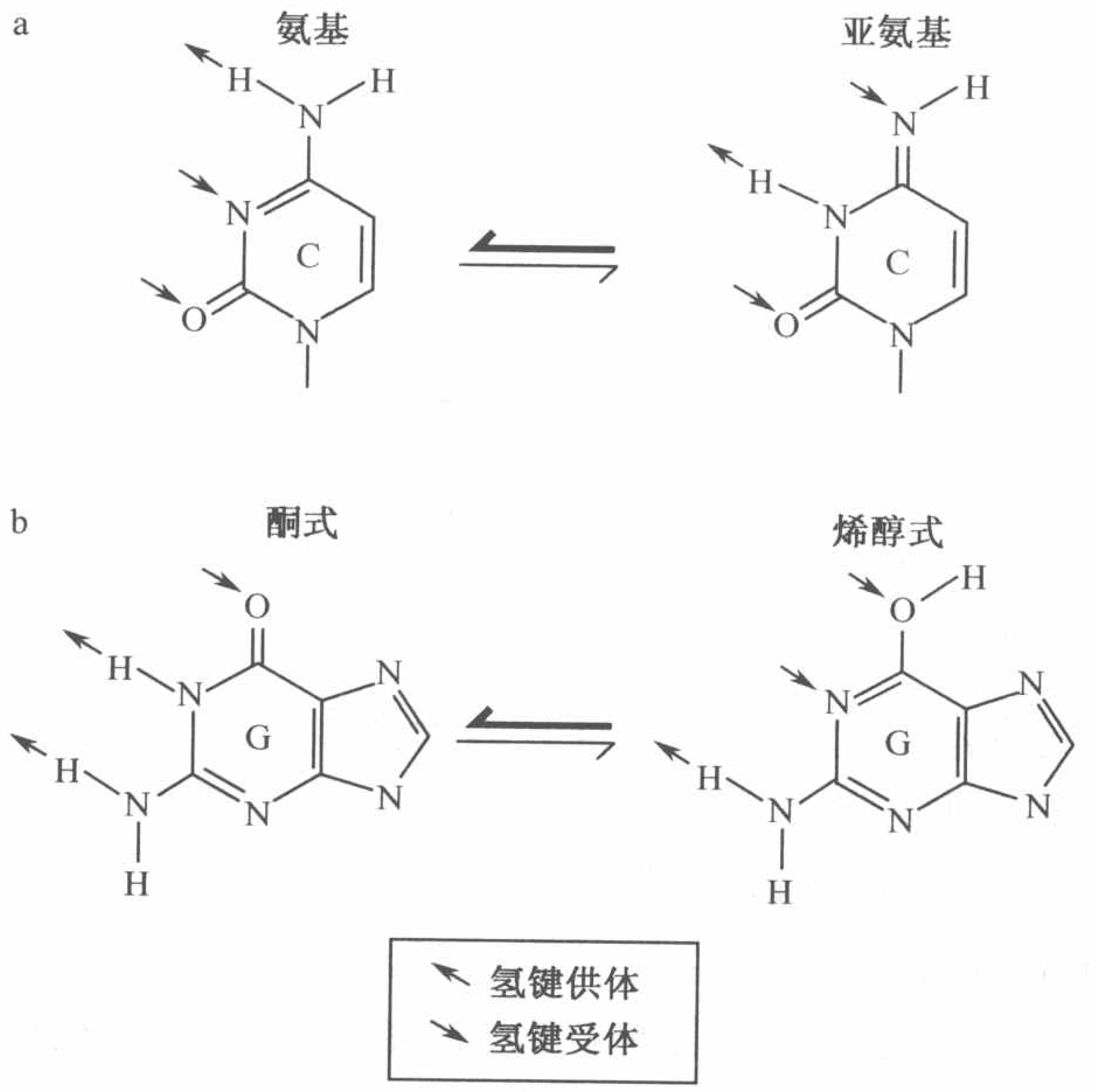


图 6-4 嘌呤和嘧啶。虚线表示碱基和核糖结合的位点。为简便起见，除非与说明有关，氢原子不在后面图示的核糖和碱基中表示出来。

图 6-5 碱基异构体。氨基⇌亚氨基和酮式⇌烯醇式的互变异构现象。a. 胞嘧啶通常为氨基式很少有亚氨基式。b. 鸟嘌呤通常为酮式很少有烯醇式。

双螺旋的两条链序列互补

腺嘌呤和胸腺嘧啶配对，鸟嘌呤和胞嘧啶配对的结果是使两条相互缠绕的链上的碱基序列呈互补关系并赋予 DNA 自我编码的特性。举个例子，如果一条链上的序列是 5′-ATGTC-3′，那么另一条链则必然是互补的序列 3′-TACAG-5′。

“Watson-Crick” 配对规则的严格性源自于腺嘌呤和胸腺嘧啶，鸟嘌呤和胞嘧啶的形状和氢键的互补性（图 6-6）。腺嘌呤和胸腺嘧啶配对，腺嘌呤 C6 上环外的氨基基团与胸腺嘧啶 C4 上的羰基可以形成一个氢键；同样，腺嘌呤的 N1 和胸腺嘧啶的 N3 也形成一个氢键。鸟嘌呤和胞嘧啶也是如此，所以在一碱基对中既有氢键也有形状的互补。GC 碱基对有 3 个氢键，因为鸟嘌呤 C2 上环化的氨基基团和胞嘧啶 C2 上的羰基形成氢键，同样鸟嘌呤的 N1 和胞嘧啶的 N3，鸟嘌呤 C6 的羰基和胞嘧啶 N4 上环化的氨

基也可形成氢键。Watson-Crick 碱基配对需要碱基处于它们优势的互变异构状态(tautomeric state)。

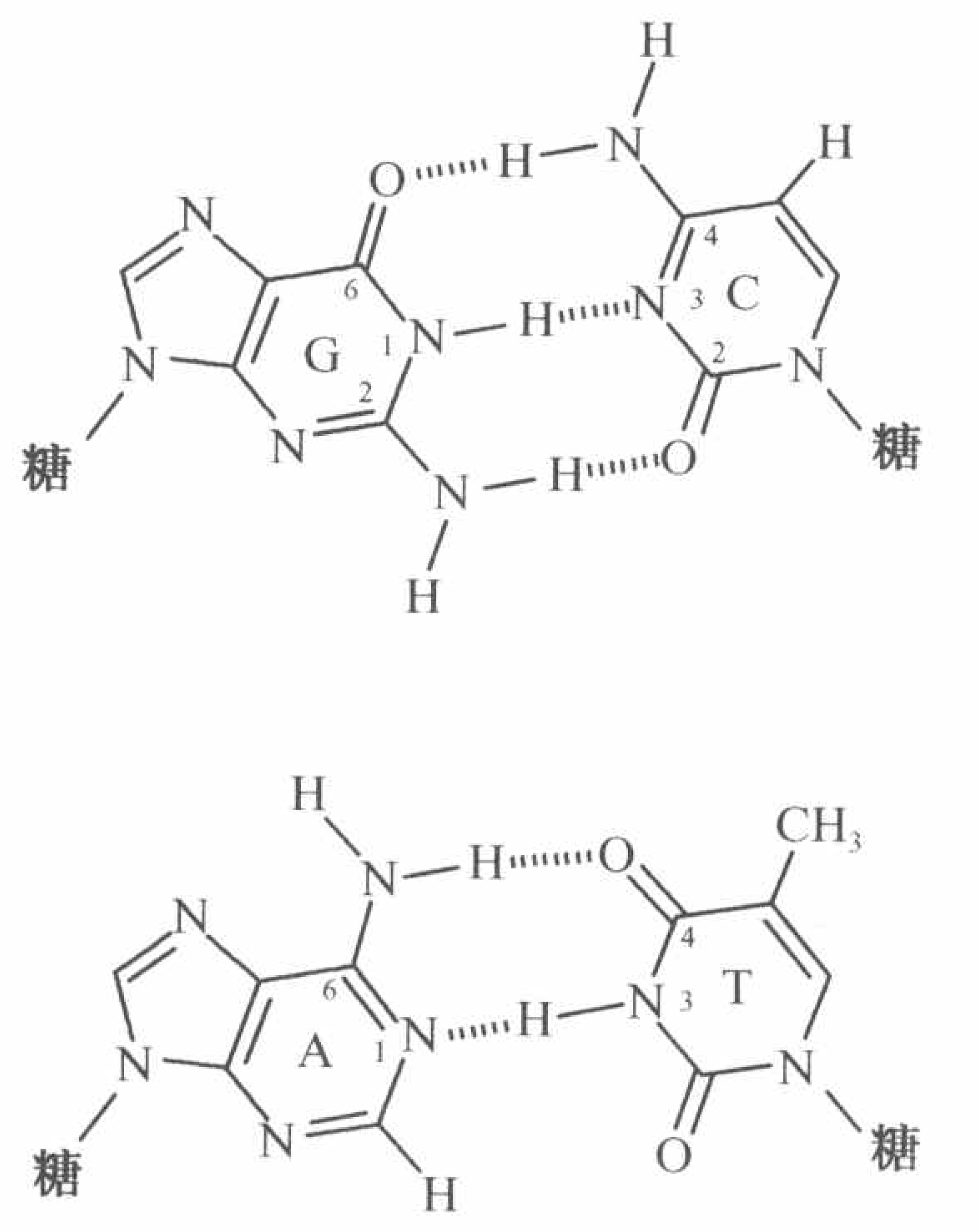


图 6-6 A:T 和 G:C 碱基配对。图示碱基之间的氢键。

双螺旋一个重要的特点是两个碱基对正好有相同的几何学结构，两个糖之间无论是 A∶T 碱基对或是 G∶C 碱基对并不会影响它们的排列，因为这两个碱基对的糖附着点间的距离是相同的。T∶A 碱基对或 G∶C 碱基对也不会影响糖的排列。换句话说，两个糖间有一个大致的二乘对称轴关系，不需要扭曲 DNA 的总体结构，这 4 种碱基对就能安置于内。而且，双螺旋的两条糖-磷酸骨架上的碱基彼此间可以整齐地堆积。

氢键对碱基配对特异性的重要作用

互补碱基之间由氢键连接是双螺旋的一个基本特征，有助于双螺旋的热动力学稳定性和碱基配对的特异性。乍一看，由于以下几个原因氢键看似对 DNA 的稳定性影响不大。水溶液中的有机分子含有满足所有氢键形成的特性，因为许多水分子可快速地与它结合或解离。结果，每当一个碱基对形成时在配对碱基之间产生一个氢键，就有一个碱基与水分子之间的氢键断裂，而这一氢键在碱基对还没形成前就存在了。因此，氢键对维持双螺旋稳定性所起的净能量作用似乎不是很大。然而，当多核苷酸链分离时，水分子在碱基上排成一排；当双螺旋中两条链合在一起时，水分子被碱基替代。这样造成了无序，增加了熵值，使双螺旋得以稳定。氢键并不是稳定双螺旋的唯一因素。第二个重要的作用力来自于碱基间的堆积力。碱基是扁平、相对难溶于水的分子，它们以大致垂直于双螺旋轴的方向互相上下堆积。双螺旋堆积时碱基之间电子云的相互作用（π-π）对双螺旋的稳定性有重要作用。

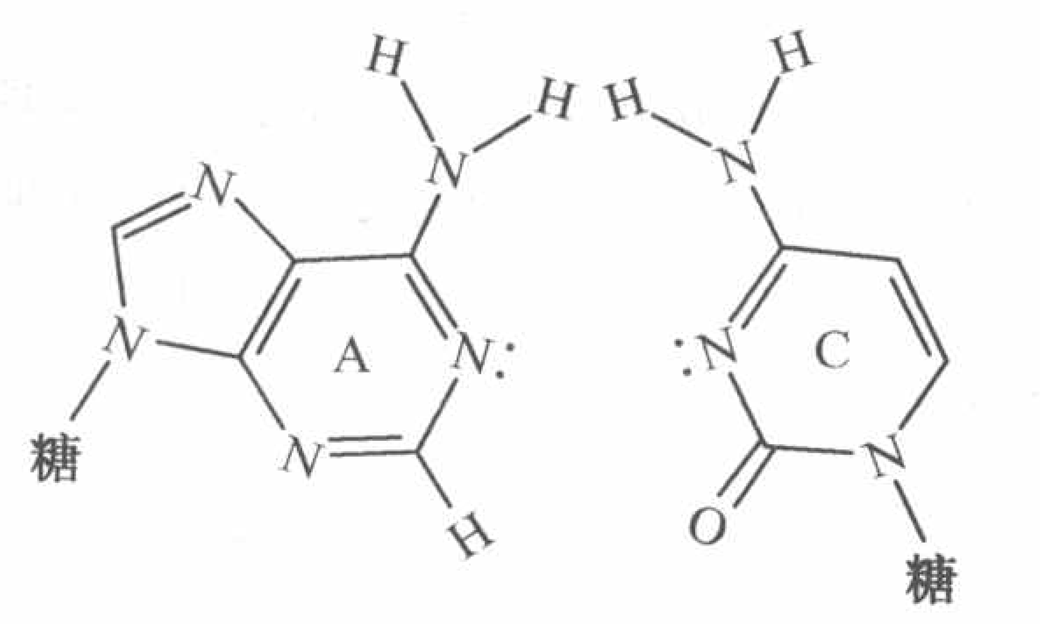


图 6-7 A:C 的不相配。本图显示 A 和 C 之间不能形成正确的氢键。因此它们间的碱基配对是不稳定的。

氢键对于碱基配对的特异性也非常重要。设想我们试着使腺嘌呤和胞嘧啶配对，这样，一个氢键受体（腺嘌呤的 N1）对着另一个氢键受体（胞嘧啶的 N3），而两个受体之间无法再容纳一个水分子，也就无法形成氢键（图 6-7）。同样，两个氢键供体——腺嘌呤 C6 和胞嘧啶 C4 上的氨基基团也彼此相对。所以，A∶C 碱基配对是不稳定的，因为受体和供体基团上的水分子必然会被除去，而碱基对间无法形成氢键。

碱基会从双螺旋中翻转出来

到目前为止我们已经知道，双螺旋的力能特征有助于使一条多核苷酸链上的一个碱基

与另一条链上的互补碱基配对。然而，有时单个的碱基会从双螺旋中突出，这种值得注意的现象叫做碱基翻出（base flipping），如图 6-8 所示。正如我们将在第 9 章中看到的，碱基从双螺旋中翻出能使碱基处于酶的催化部位，一些酶就是通过和这些螺旋构型外的碱基相互作用而使碱基甲基化或除去受损碱基的。而且，人们相信与同源重组和 DNA 修复相关的酶能将碱基一个接一个地翻转出来，借此扫描 DNA，搜寻同源区域或损伤区域。因为一次只翻转一个碱基，这个过程并不需要很多能量。显然，DNA 比人们所设想的更有弹性。

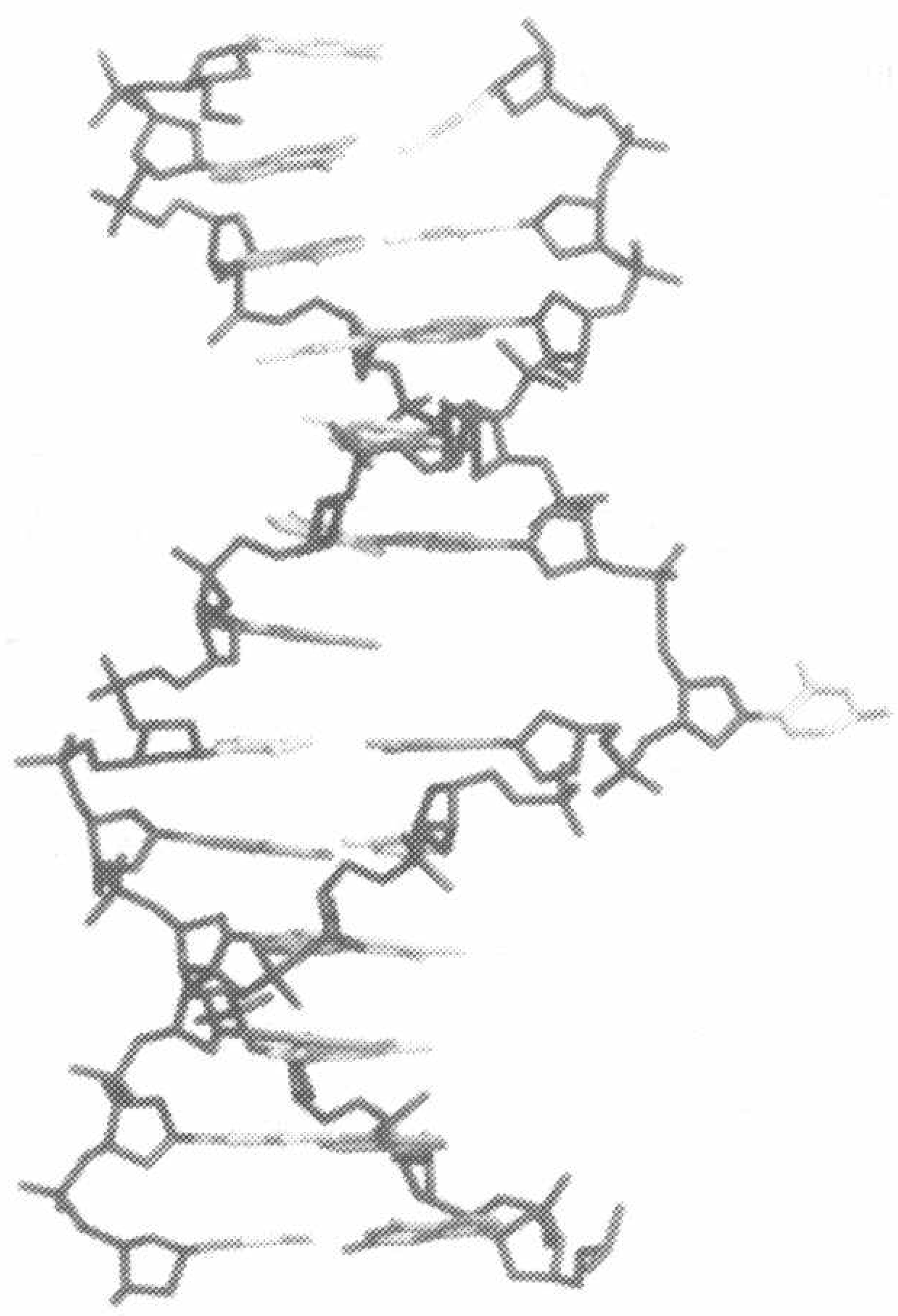


图 6-8 碱基翻出。图中的 DNA 结构由甲基化酶的结构中分离得到，显示了胞嘧啶残基的翻出以及给邻近碱基对造成的轻微变形。(Klimasauskas S., Kumar S., Roberts R. J., and Cheng X. 1994. Cell. 76：357，图像由 BobScript、MolScript 及 Raster 3D 软件制作。)

DNA 通常是右手双螺旋

用物理学中的手性法则进行分析，我们可以看到双螺旋中两条多核苷酸链都是右手螺旋。试想一下，举起你的右手如图 6-9 所示的 DNA 分子，拇指朝上并沿着螺旋的长轴，其余手指随着沟的方向摆放。沿拇指所指的方向追踪双螺旋的一条链，你将注意到螺旋的方向和你的手指所指的方向相同。若用左手，则无法得到上述发现。

DNA 的螺旋状结构导致了它是有周期性的。两个相邻碱基对间的夹角大约是 36°。通过分析 DNA 的 X 射线晶体衍射图发现，完成一个螺旋（360°）需 10 个碱基对的堆积（图 6-1a），也就是说，每一个螺周通常是 10 个碱基对。进一步的讨论请看框 6-1，云母实验表明溶液中双螺旋 DNA 每一螺周含有 10.5 个碱基对。

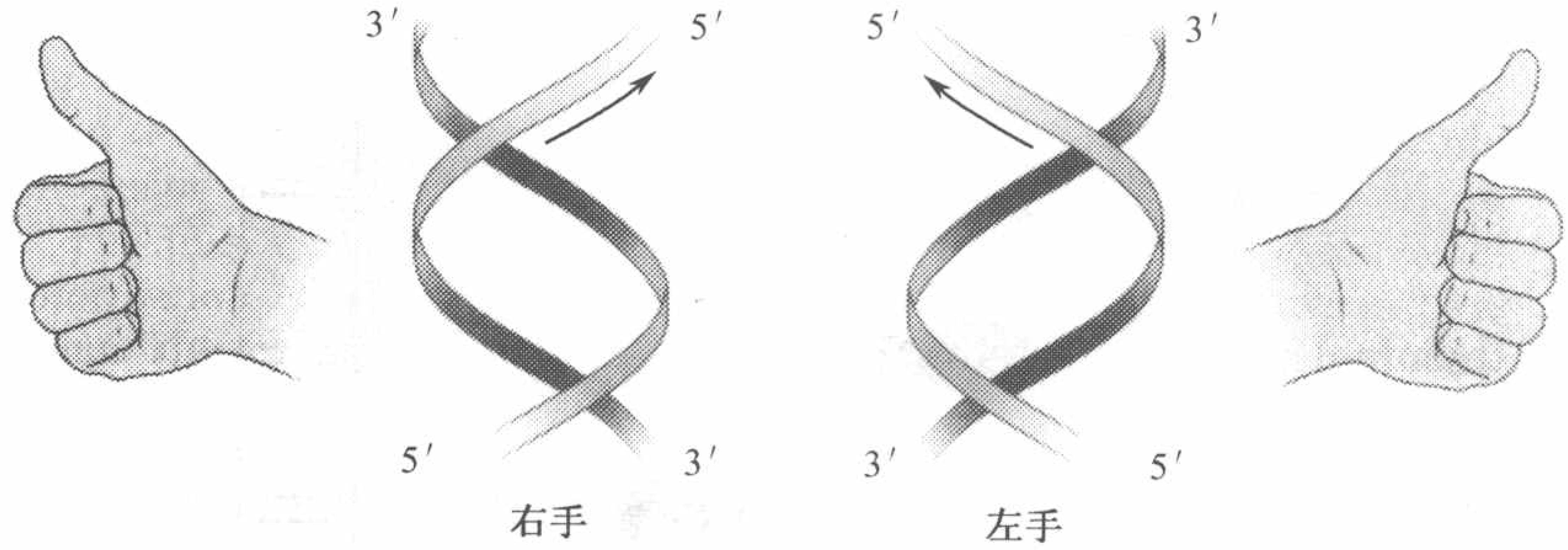


图 6-9 左手和右手螺旋。双螺旋的两条链以右手螺旋的方式互相缠绕。

双螺旋有大沟和小沟

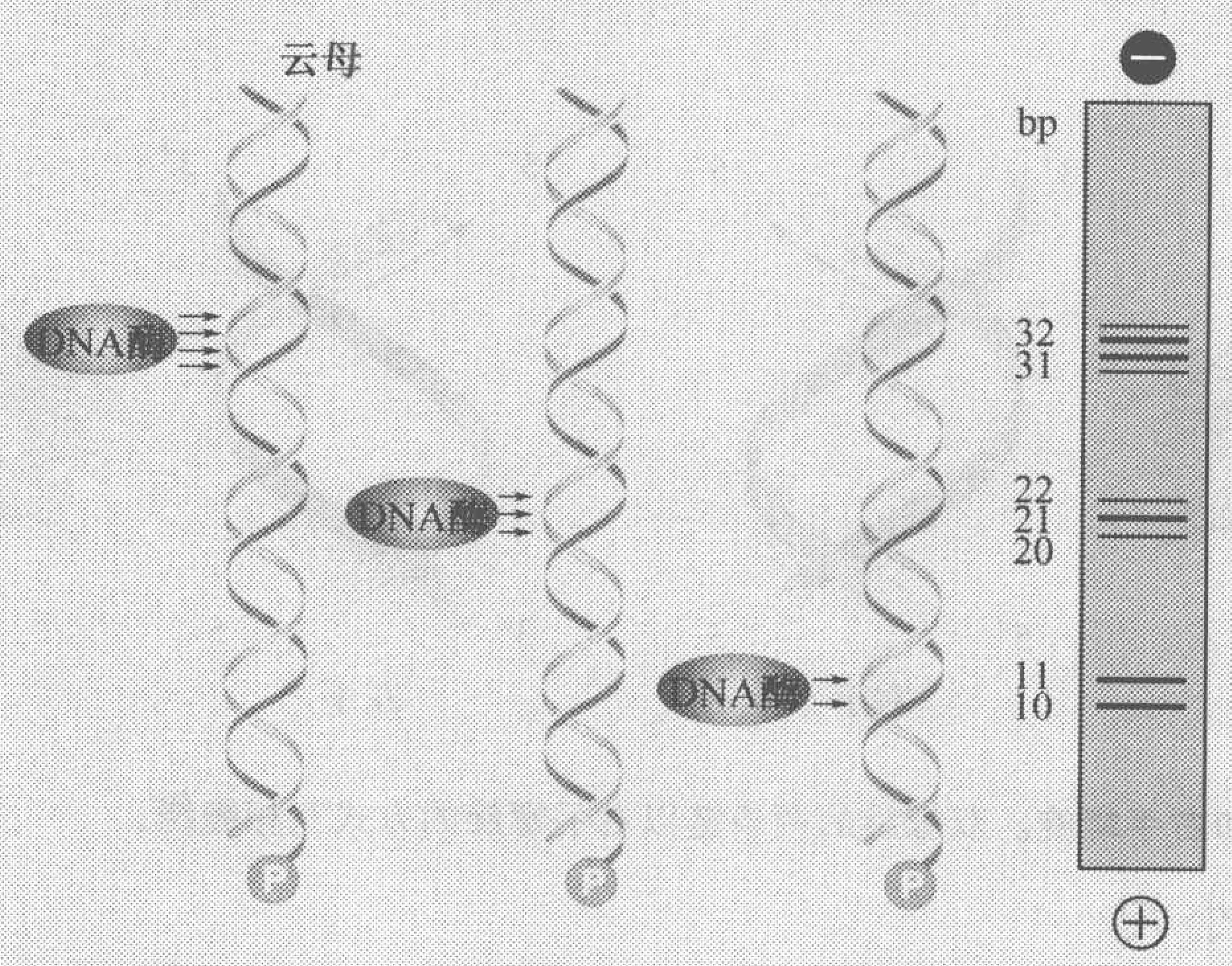
DNA 分子两条链所形成的双螺旋结构使 DNA 分子延伸成长长的多聚体。其上有两条大小不一的沟。为什么会有一条小沟一条大沟？这是由碱基对的空间几何结构决定

的。碱基对上凸起的两个糖（糖苷键间的夹角）之间的角度大约为 120°（指窄角，广角则为 240°）（图 6-1b 和图 6-6）。结果，当越来越多的碱基对上下堆积起来时，其一侧糖间的窄角就形成了小沟（minor groove），而另一侧的广角就形成了大沟（major groove）。（如果糖和糖之间呈直线相对，也就是 180°的角度，那么所形成的两个沟大小相同，不存在小沟和大沟之分。）

框 6-1 云母实验表明溶液中双螺旋 DNA 每一螺周含有 10.5 个碱基对

每一螺周含有 10 个碱基对，这个数值会因为条件的不同而有所变动。20 世纪 70 年代进行的经典实验表明吸附在某一表面上的 DNA 每一螺周不只含有 10 个碱基对。小片段 DNA 可以吸附在云母表面上，且 5′端磷酸基的存在使得 DNA 在云母上呈一定的方向性。随后吸附在云母上的 DNA 用 DNase Ⅰ（一种脱氧核糖核酸酶）消化。这种酶可以裂解 DNA 磷酸-糖骨架上的磷酸二酯键。但由于这种酶的体积大，空间阻碍使它无法作用于 DNA 的侧面或底面，仅仅可以切断距离云母最远的 DNA 表面上的磷酸二酯键（将 DNA 看作平躺于平面上的圆柱体）。因此经过切割后得到的片段的长度应该可以反映出 DNA 的周期性，即每一螺周碱基对的数目。

吸附在云母上的 DNA 经过 DNase Ⅰ切割后得到的片段可以通过聚丙烯酰胺凝胶电泳进行分离（框 6-1 图 1，第 20 章关于电泳的说明）。聚丙烯酰胺凝胶是一种类似于果冻的基质。因为 DNA 带负电，所以它在胶体中向电场的正极泳动。由于胶体的阻碍作用，使得片段的泳动速度与片段的大小成反比，即片段越大泳动速度越慢。电泳分离后，从电泳结果中我们发现片段平均大小为 10 和 11、21、31 和 32 等，即以 10.5 个碱基对逐倍递增的 DNA 片段成簇存在。10.5 就是每一螺周的碱基对的数目，这与其他方法推测出的溶液中 DNA 每一螺周的碱基对数目相近（见下面“双螺旋以多重构象存在”）。利用 DNase 探测 DNA 结构的这种方法现已应用到分析 DNA 与蛋白质的相互作用上（第 17 章）。



框 6-1 图 1 云母实验

大沟中有丰富的化学信息

每个碱基对的边缘都暴露于大沟、小沟中，从而产生了氢键供体和受体及识别碱基对的范德华力表面的模式（图 6-10）。在大沟中的 A∶T 碱基对边缘，化学基团以如下顺序排列：氢键受体（腺嘌呤 N7）、氢键供体（腺嘌呤 C6 上的环外氨基）、氢键受体（胸腺嘧啶 C4 上的羰基）和大的疏水表面（胸腺嘧啶 C5 上的甲基）。同样在大沟中 G∶C碱基对的边缘展示了下列基团：氢键受体（鸟嘌呤 N7）、氢键受体（鸟嘌呤 C6 上的羰基）、氢键供体（胞嘧啶 C4 上的环外氨基）和小的非极性氢（胞嘧啶 C5 上的氢）。

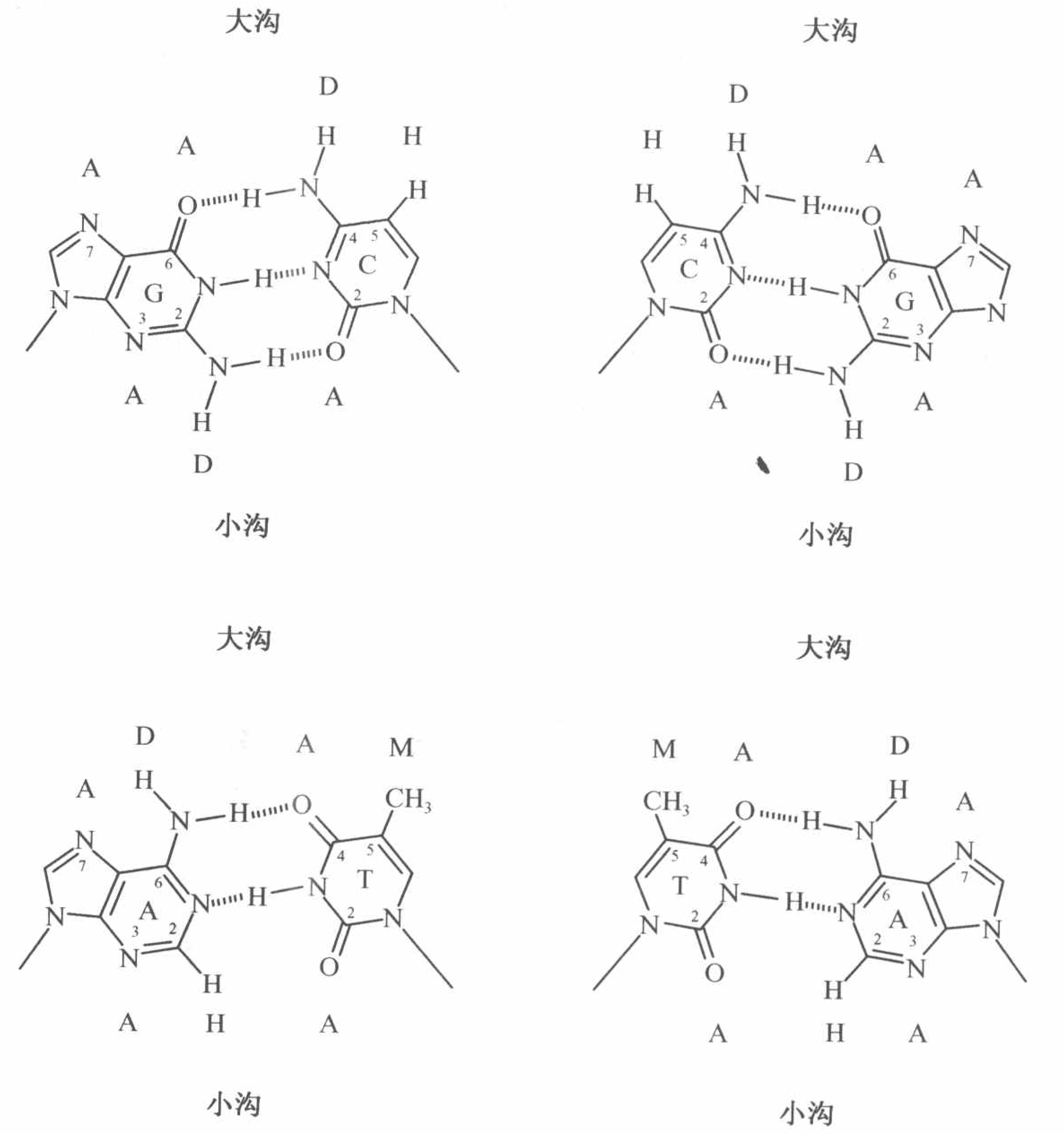


图 6-10　碱基对边缘的化学基团暴露于大沟、小沟中。红色字母分别为氢键受体（A）、氢键供体（D）、非极性氢（H）和甲基基因（M）。

因此，所形成的氢键和暴露于大沟中的整体形状的特有模式可以将 A∶T 碱基对与 G∶C 碱基对区别开来，并可区分 A∶T 碱基对与 T∶A 碱基对、G∶C 碱基对与 C∶G 碱基对。我们可以把这些特征看作代码，其中 A 代表氢键受体，D 代表氢键供体，M 代表甲基，H 代表非极性氢。以这种代码方式，在大沟中 ADAM 代表一个 A∶T 碱基对；AADH 代表一个 G∶C 碱基对；MADA 代表一个 T∶A 碱基对；HDAA 代表一个 C∶G 碱基对。在所有情况下，大沟中化学基团的这种代码都可以具体说明不同碱基对的特性。这些模式是非常重要的，使得蛋白质无需解开继而破坏双螺旋就可准确地识别

DNA 序列。事实上，正如我们将要看到的，一个首要的解码机制就是有赖于氨基酸侧链进入大沟、识别特定 DNA 并与其结合的能力。

小沟中化学信息较少，而且现有的信息对于区分各碱基对也没多大用处。由于体积小它不太能容纳氨基酸侧链。在小沟中，A∶T 碱基对与 T∶A 碱基对、G∶C 碱基对与C∶G碱基对看起来极其相似。A∶T 碱基对有氢键受体（腺嘌呤 N3）、非极性氢（腺嘌呤 C2 上的氢）和氢键受体（胸腺嘧啶 C2 上的羰基），它的代码即为 AHA，但若从反向读取，其代码是相同的。从侧链伸入小沟的蛋白质的氢键形成特性角度看，A∶T碱基对与 T∶A 碱基对没什么不同。同样，G∶C 碱基对有氢键受体（鸟嘌呤 N3）、氢键供体（鸟嘌呤 C2 上的环外氨基）和氢键受体（胞嘧啶 C2 上的酮基），它的代码为 ADA。所以，从氢键形成角度看，C∶G 碱基对与 G∶C 碱基也没有什么不同。小沟中，A∶T 碱基对与 G∶C 碱基对容易区分，但 G∶C 碱基对与 C∶G 碱基对、A∶T碱基对与 T∶A 碱基对则不容易区分（图 6-10）。

双螺旋以多重构象存在

早期采用高浓度的抽成细纤维的 DNA 溶液对 DNA 进行 X 射线晶体衍射分析，提出了两种结构：A-DNA 和 B-DNA（图 6-11；框 6-2：如何通过 X 射线胶片中的斑点来

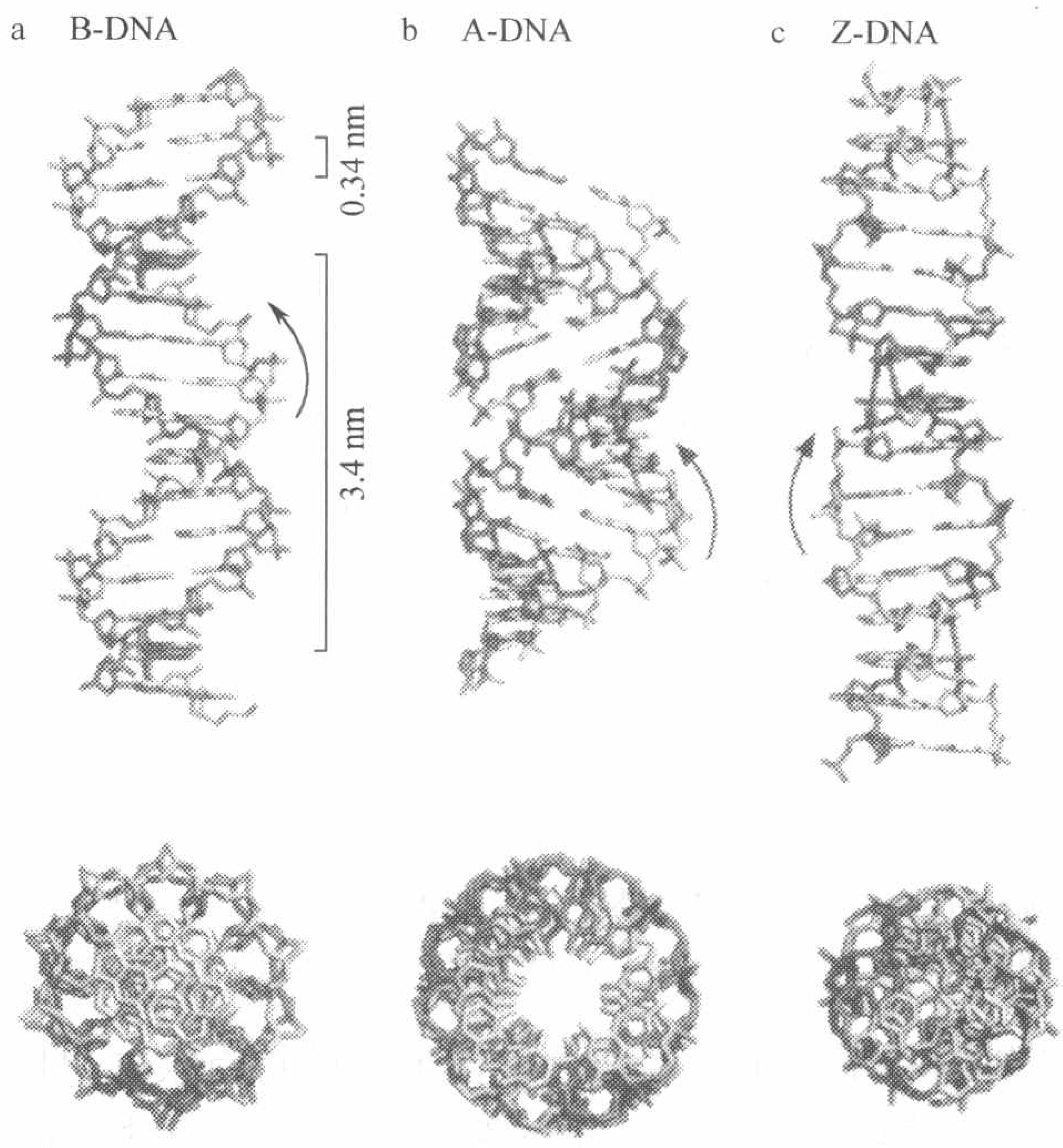


图 6-11　B-DNA、A-DNA 和 Z-DNA 的模型。所有结构中每条链的糖-磷酸骨架在外侧（一个用紫色表示，一个用绿色表示），碱基在内侧（用银色表示）。上面的为侧面图，下面的为横切图。a. B-DNA 是细胞中的通常构型，其特征为每螺周 10 个碱基对（螺距 3.4nm），即邻近碱基对堆积距离为 0.34nm，并且可以看到大沟和小沟。b. A-DNA 更为紧凑，每螺周 11 个碱基对，且碱基对与螺旋纵轴间的夹角比较大。此外，A-DNA 是中空的（见横切图），RNA 双链和 RNA-DNA 杂交双链采取这种构型。c. Z-DNA 是左手螺旋，呈 Z 字型。（来源：C. Kielkopf and P. B. Dervan 惠赠）

揭示 DNA 结构)。在较高湿度下观察到的 B-DNA 最接近生理条件下 DNA 的一般结构:每螺周 10 个碱基对,宽的大沟和窄的小沟。在较低湿度条件下观察到的 A-DNA 每圈 11 个碱基对。与 B-DNA 相比,大沟窄且深,小沟宽且浅。细胞中绝大部分 DNA 以 B 型构象存在,但在某些 DNA-蛋白质复合体中 DNA 也可以 A 型构象存在。而且 A-DNA 与以双螺旋形式存在的 RNA 分子的结构相似。

重要实验

框 6-2 如何通过 X 射线胶片中的斑点揭示 DNA 结构

在分子生物学的历史上最经典的图像之一,就是由 Rosalind Franklin 拍摄的一张著名的照片:一个定向的 DNA 分子纤维的 X 射线衍射模式图。这张由她拍摄的照片之所以具有如此重大的历史意义,是因为该照片为 B 型 DNA 的沃森-克里克模型(Waston-Crick model)提供了关键的证据。而且,Francis Crick 作为一位推动了螺旋分子衍射理论发展的科学家,从胶片上的斑点的模式推断出 DNA 的两条链是相互缠绕的。乍一看,她的图像与双螺旋并没有多大关系。那么,这些斑点的神秘模式是如何揭示出作为遗传物质的 DNA 的原子结构呢?

从图中我们可以看到,Franklin 的图像出一个中心 Maltese 十字视标(Maltese cross)构成(在框 6-2 图 1 中用红色表示),该视标又由宽点(该宽点的宽度反映出 DNA 分子纤维的无序状态)组成。斑点沿着水平“层次”(layer)线(见数字标注)被均匀隔开。请注意,从十字视标的中心往上或往下计数时,不能发现第四层次线上的斑点。同时还要注意,Maltese 十字视标以及图像顶部和底部极暗的区域产生出一系列四菱形(four diamond-shaped)的区域(见用蓝色表示的两个例子)。经过对波的衍射本性的一些思考,从定性的角度上,我们现在可以这样解释,这个表面上晦涩难懂的斑点模式其实与双螺旋结构是相对应的。

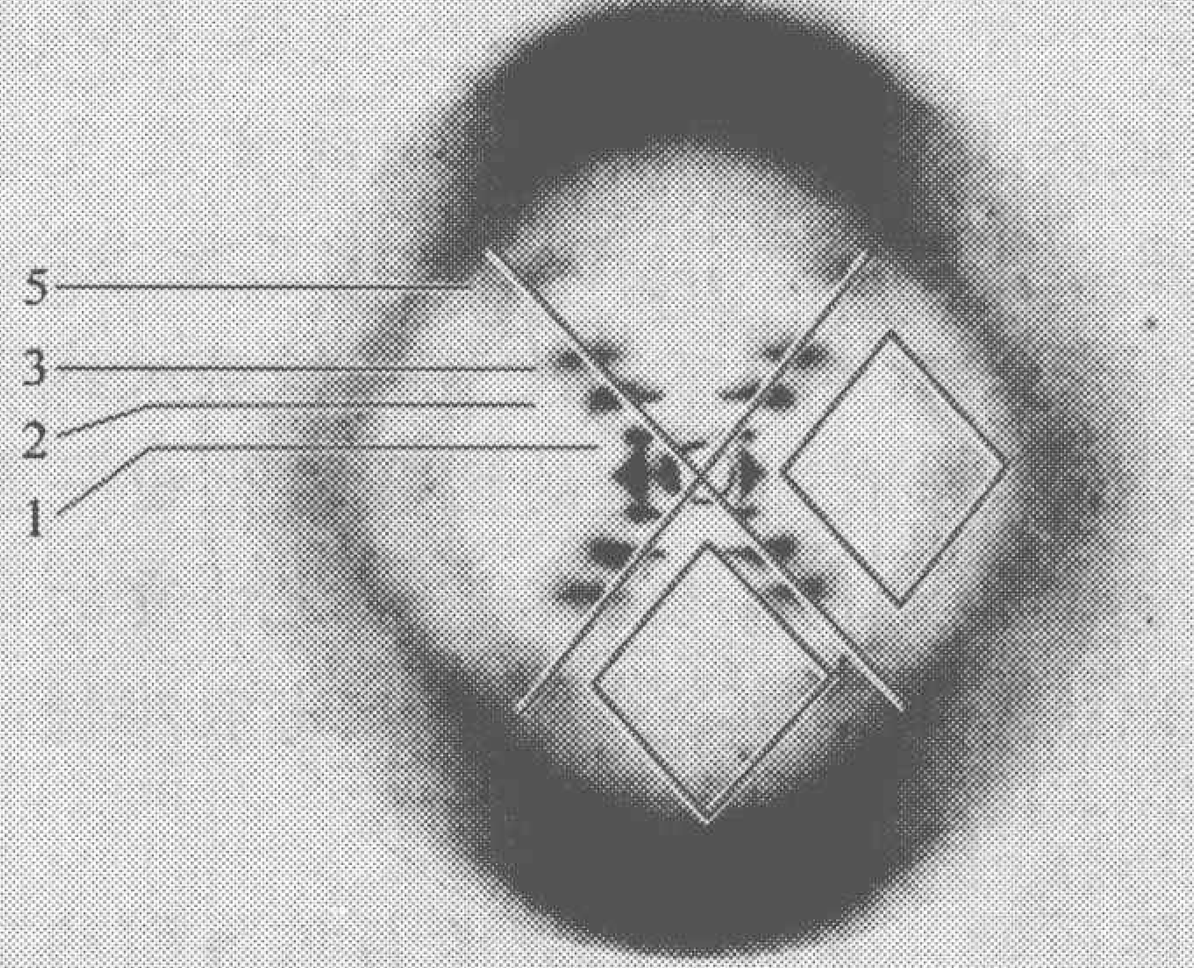
X 射线衍射的原理是当波通过一个周期性的阵列时,如果波的波长与该阵列的重复单位近似的话,波之间会发生干涉(interference)(这就是为什么 X 射线会被用来揭示原子结构,因为它的波长很短,仅为 0.15nm)。如果波的振动一致,波之间会相互加强(结构性干涉),但如果一系列波的波谷与另一系列波的波峰相遇,波之间则相互抵消(破坏性干涉)。因此,一束束通过含有水平线的阵列的波会产生一行垂直于该水平线的斑点(框 6-2 图 2)。可以想象,如果水平线倾斜的话,会产生一行倾斜的斑点(该行斑点仍然垂直于倾斜的线)。可以进一步想象,对于如图所示的两套以“Z”字形连接的倾斜线来说,如果波通过它们,则会产生由两行倾斜的斑点组成的十字形。

现在让我们回到 DNA 上来。想象一下双螺旋中一条链的骨架投影到一个平面上来,不严格地说,这会产生一系列相互连接的 zig 和 zag(或更恰当一点,正弦曲线)。如果我们这样,将 zig 想像为产生一套倾斜线,zag 产生另一套倾斜线,那么穿过 zig 和 zag 的波将会产生如以上例子所示的两行相互交叉的斑点。这就构成了 Franklin 照片中 Maltese 十字视标的基础,因此该十字揭示出了 DNA 的螺旋结构。基于 X 射线的波长,

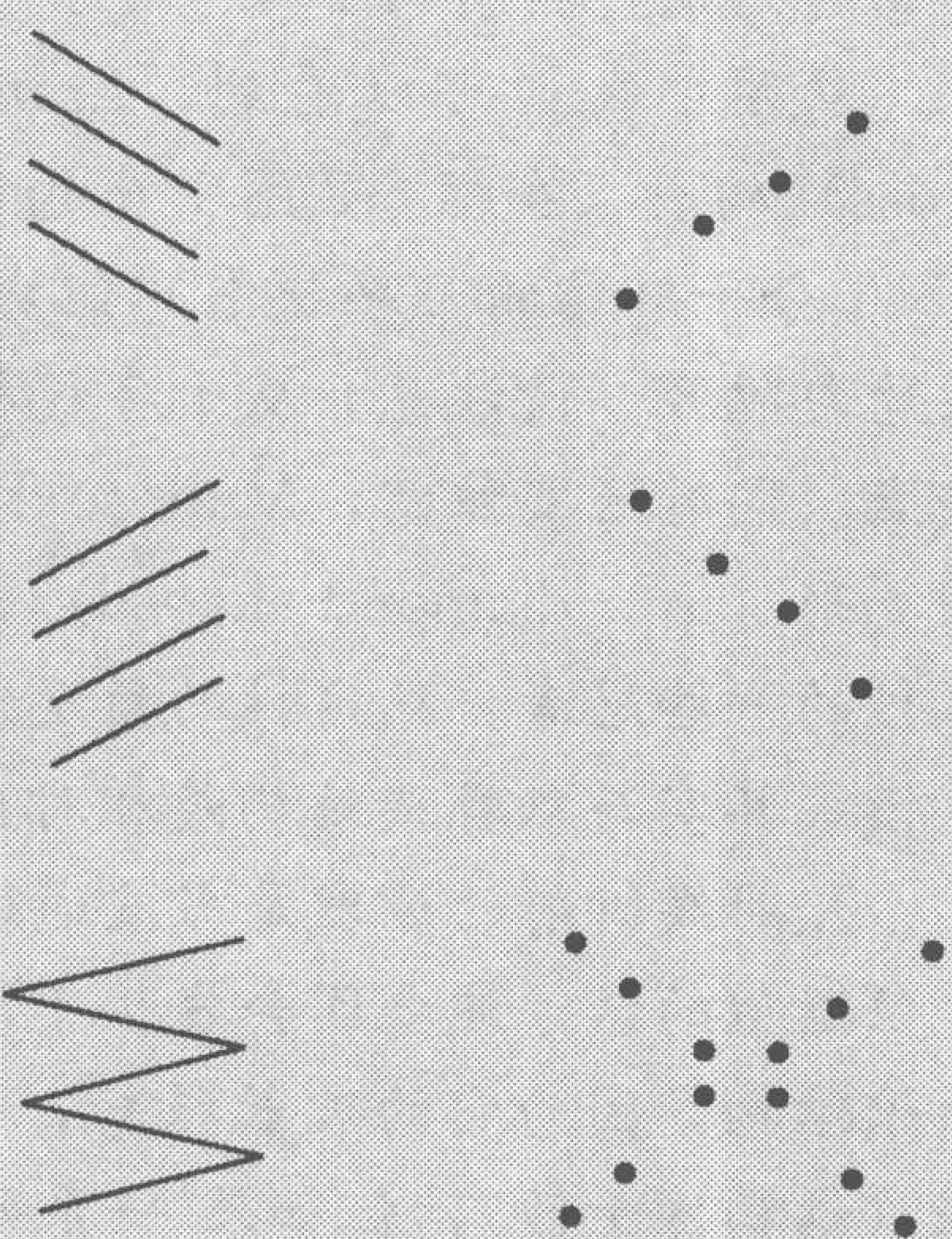
同时测量层次线间隔开的距离，进一步揭示出螺旋的周期为 3.4nm。当然，DNA 并非只包含一个螺旋骨架，而是包含双螺旋骨架，这一点也通过 Fraklin 的照片得以揭示。DNA 双螺旋之间有 3/8 的螺旋重复单元对位不准，结果证明双螺旋之间的这种抵消产生了一个额外的破坏性干涉，而将第四层次线抹除。因此，丢失的第四层次线表明 DNA 具有双螺旋结构，并告诉我们双螺旋之间是如何相对排列的。

最后，DNA 骨架并非我们想象中的那样是一条平滑线。相反，从原子水平来讲，它是颗粒性的，由糖-磷酸单元构成。这种颗粒性导致了额外的强度，尤其是十字视标中心的北面和南面，产生出一种四菱形的模式。在比此处所示图片具有较高分辨率的照片中，我们从十字视标中心往北面和南面可以清楚地数出 10 条层次线。这种衍射模式的特点告诉我们，双螺旋的周期（3.4nm）为原子周期的 10 倍，对应于以 0.34nm 为间隔的十个重复单元。因为每一个糖-磷酸单元具有一个碱基，因此 B 型 DNA 每一螺旋周期（即每一个螺旋转角）由 10 个碱基对组成。

因此，对于波的衍射的初步理解使如下工作成为可能：通过 X 射线胶片上的斑点的简单模式推导出 DNA 结构的主要特点。



框 6-2 图 1 Rosalind Franklin 的 DNA 的 X 射线衍射成像揭示出 Maltese 十字视标（Maltese cross）。（经 Franklin R. E. and Gosling R. G. 1953. *Nature* 171：740～741. © Macmillan 许可修改）



框 6-2 图 2 穿过平行线的波的衍射模式

B-DNA 是一种理想结构，与细胞中的 DNA 略有不同，体现在两方面：第一，溶液中的 DNA 分子比 B-DNA 分子螺旋程度更高，平均每螺周有 10.5 个碱基对；第二，B 型构象是均一的结构，然而实际的 DNA 没有如此规则，甚至从精细结构看，它的各个碱基对之间都有所不同，这是将不同序列 DNA 的晶体结构进行比较得出的结论。例如，每个碱基对的两个碱基并不是一直处于同一平面，甚至能呈现“螺旋桨扭转”，两个碱基平面沿着碱基对的长轴相对于彼此向相反方向旋转，赋予碱基对类似螺旋桨的特征（图 6-12）。此外，每个碱基对精确的螺旋桨扭角并不恒定，结果局部上大沟小沟的宽度会发生变化。因此，DNA 分子永远不会是完全规则的双螺旋。相反，DNA 的准确构象取决于双螺旋每一位置上何种碱基对（A∶T、T∶A、G∶C 或者 C∶G）及其相邻碱基对的存

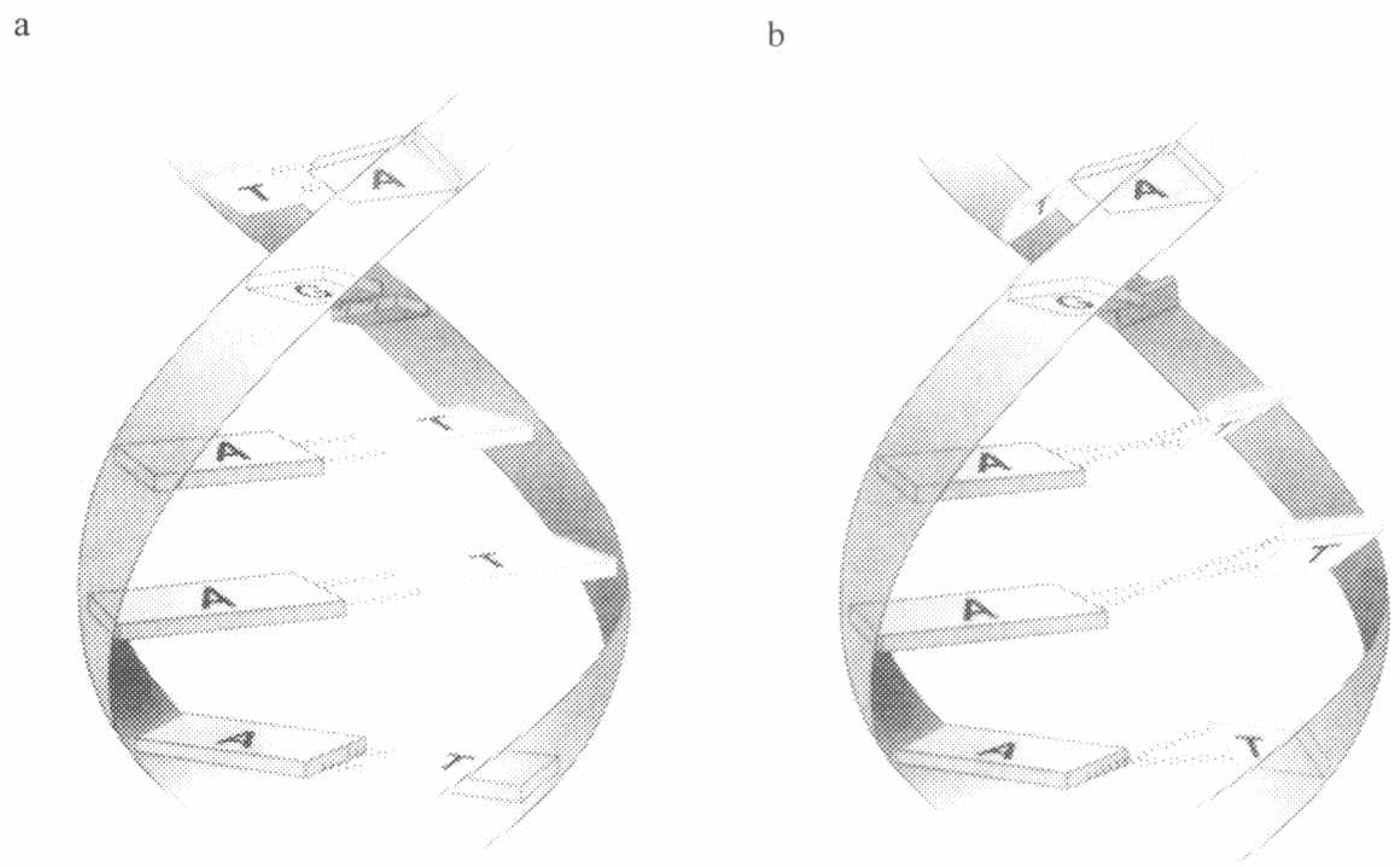


图 6-12 右手螺旋中嘌呤和嘧啶碱基对之间的螺旋桨扭转。 a. 常规 Watson 和 Crick 配对中 3 个连续 A：T 碱基对的序列结构。b. 螺旋桨扭转引发碱基相对于其长轴的旋转。（据 Aggaarwal et al.，1988. Science 242：899-907，Fig. 5b. © AAAS.）

在。但是就许多研究而言，B 型构象仍然是与细胞中 DNA 结构最接近的。

DNA 有时可形成左手螺旋

包含交替的嘌呤和嘧啶残基的 DNA 除可形成右手螺旋外还可形成左手螺旋。为了弄清左手螺旋是如何形成的，我们需要仔细考虑将碱基连到 2′-脱氧核糖 1′位置的糖苷键。这个键可以以顺式（*syn*）或反式（*anti*）两种构象中的一种存在（图 6-13）。右手螺旋 DNA 的糖苷键只以反式构象存在。而左手螺旋中基本重复单位通常是嘌呤-嘧啶二核苷酸。糖苷键在嘌呤位置以顺式构象存在，在嘧啶位置以反式构象存在。正是嘌呤处的顺式构象造成了左手螺旋。由于嘌呤处变为顺式构象使得左手螺旋 DNA 的骨架上顺式反式构象交替出现，呈现 Z 字型走向（因此命名为 Z-DNA；图 6-11），可与右手螺旋区分开。影响反式向顺式转变的旋转也使核糖基团在折叠过程中发生了变化。如图 6-13所示，C3′和 C2′可以转换位置。只有在高阳离子浓度下（如 Na^+），中和了带负电的磷酸基团，溶液中的交替嘌呤-嘧啶残基才会呈现左手螺旋构象。在低盐浓度情况下，它们会形成典型的右手螺旋构象。Z-DNA 的生理意义还不清楚，左手螺旋在细胞 DNA 中可能只占很小一部分。A-DNA、B-DNA 和 Z-DNA 的更多细节列于表 6-2。

表 6-2 从 X 射线单晶体衍射分析得出的 A 型、B 型和 Z 型 DNA 结构特性的比较

	螺旋类型		
	A	B	Z
外形	短粗	适中	细长
每碱基上升	23Å	3.32Å	3.8Å
螺旋直径	25.5Å	23.7Å	18.4Å

续表

	螺旋类型		
	A	B	Z
螺旋方向	右手	右手	左手
每螺旋单位碱基数	1	1	2
每螺周碱基数	约 11	约 10	12
碱基夹角	33.6°	35.9°	每 2bP 为—60°
螺距	24.6Å	33.2Å	45.6Å
碱基倾角	+19°	—1.2°	—9°
碱基平均螺旋桨扭角	+18°	+16°	约 0°
螺旋轴位置	大沟	穿过碱基对	小沟
大沟	极狭、很深	宽、中等深度	平坦
小沟	很宽、浅	狭、中等深度	极狭、很深
糖苷键构象	反式	反式	C 反式、G 顺式

来源：据 Dickerson R. E. et al. 1982. CSHSQB47：14. Copyright © 1982 Cold Spring Harbor Laboratory Press. 许可修改。

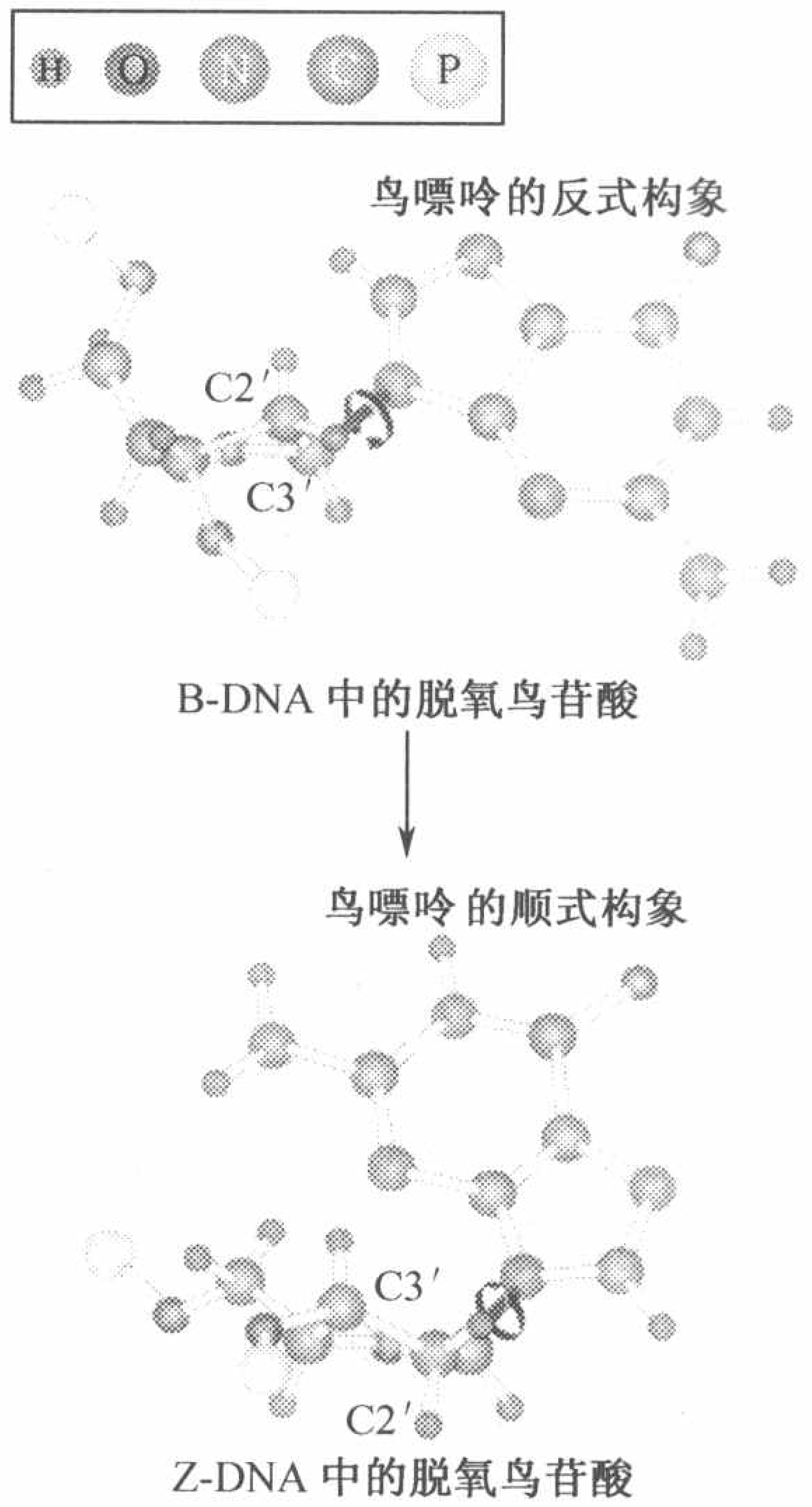


图 6-13 B-DNA 中鸟嘌呤的反式构象和 Z-DNA 中鸟嘌呤的顺式构象。右手螺旋的 B-DNA 中，连接脱氧核糖和碱基的糖苷键（用红色表示）只以反式构象存在；左手螺旋的 Z-DNA 中，糖苷键沿箭头方向旋转，在嘌呤处形成顺式构象（这里为鸟嘌呤）；而嘧啶处依然保持规则的反式构象，没有发生旋转。（来源：Wang A. J. H. et al. 1982. CSHSQB 47：41. Copyright© 1982 Cold Spring Harbor Laboratory Press.）

DNA 双链可以分开（变性）和复性

因为双螺旋的两条链依靠较弱的力（非共价键）结合在一起，故我们可预计它们分开也很容易。确实，双螺旋的结构解析之时，就表明 DNA 就以这种方式进行复制。当 DNA 溶液温度高于生理温度（接近 100℃）或者 pH 较高时，互补的两条链就可分开，这个过程称为变性（denaturation）。然而，DNA 双链完全分开的变性过程是可逆的，当变性 DNA 的热溶液缓慢降温，DNA 的互补链又可重新聚合，形成规则的双螺旋（图 6-14）。由于变性的 DNA 分子具有复性的能力，因此来源不同的两条 DNA 链通过缓慢降温也可形成人工杂交 DNA 分子，同样，互补的 DNA 和 RNA 链也可形成杂交分子。在第 20 章里，两条单链核酸形成杂交分子的能力被称为杂交（hybridization）。这是分子生物学中几项不可缺少的技术的基础。例如，Southern 印迹杂交（第 21 章）和 DNA 芯片分析（第 19 章，框 19-1）。

20 世纪 50 年代通过在一系列条件下研究 DNA 变性的经典实验，对双螺旋的特性有了深刻的认识。这些实验依靠测量通过

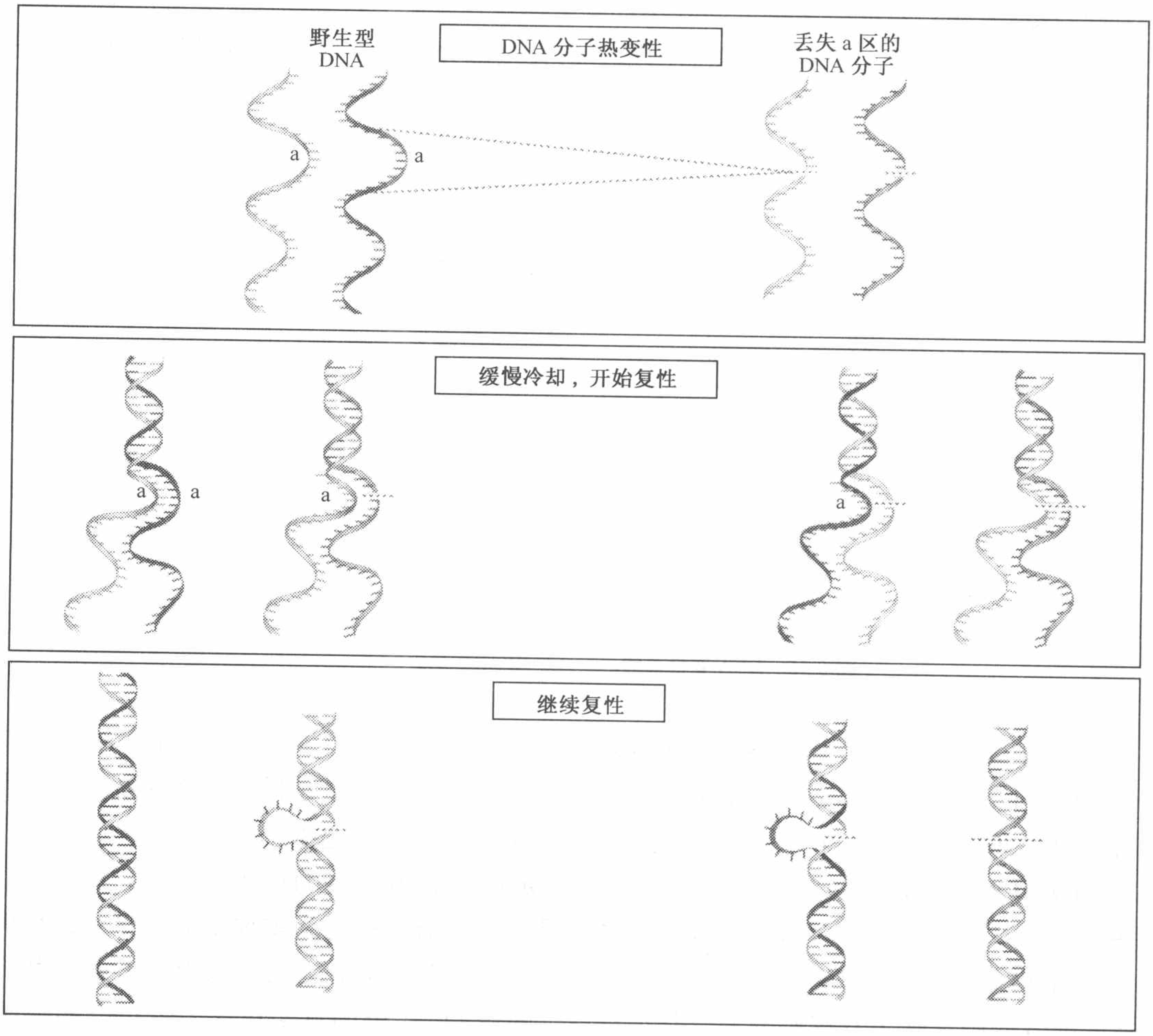


图 6-14　退火和杂交。两个基本相同的双链 DNA 分子的混合物：一个是正常的野生型 DNA，另一个是缺少一小段核苷酸的突变体 DNA（缺少核苷酸的区域用红色标为 a）。加热混合物使 DNA 变性。变性 DNA 在低于熔解温度的条件下培育时可以复性，产生两种类型的复性分子：一是完全复性分子，分别由互补的两条野生型链或两条突变型链构成双螺旋；二是杂交分子，由一条野生型链和一条突变型链构成，并形成一段短的不配对的环（区域 a）。

DNA 溶液的紫外线吸光率实现对 DNA 变性的监控。DNA 在 260nm 处有最大吸收，碱基是造成吸收的主要因素。由于双螺旋 DNA 的光吸收比单链 DNA 少 40%。当 DNA 溶液温度升高到接近水的沸点时，光吸收值（absorbance）在 260nm 处明显增加，这种现象称为增色效应（hyperchromicity）；而减色效应则是因为双螺旋 DNA 中的碱基堆积降低了对紫外线的吸收能力。

如果我们将 DNA 的光吸收对温度绘制函数曲线，不难看出光吸收的急剧增加发生在相对较窄的一个范围内。吸光度增加到最大值一半时的温度叫做 DNA 的熔点（melting point），用 T_m 表示（图 6-15）。DNA 像冰一样熔解，经历了从高度有序的双螺旋结构向无规则单链结构的转变。在 DNA 熔点处光吸收的急剧增加说明 DNA 互补双链的变性、复性是一个高度协同的、类似于拉拉链一样的过程。例如，复性最可能以缓慢

的成核反应发生，首先一条链上的一小部分碱基在互补链上找到与它们互补的那部分碱基并与之配对（图 6-14 中图），然后两条链的其余部分就会在成核位置像拉拉链那样迅速形成双螺旋（图 6-14 下图）。

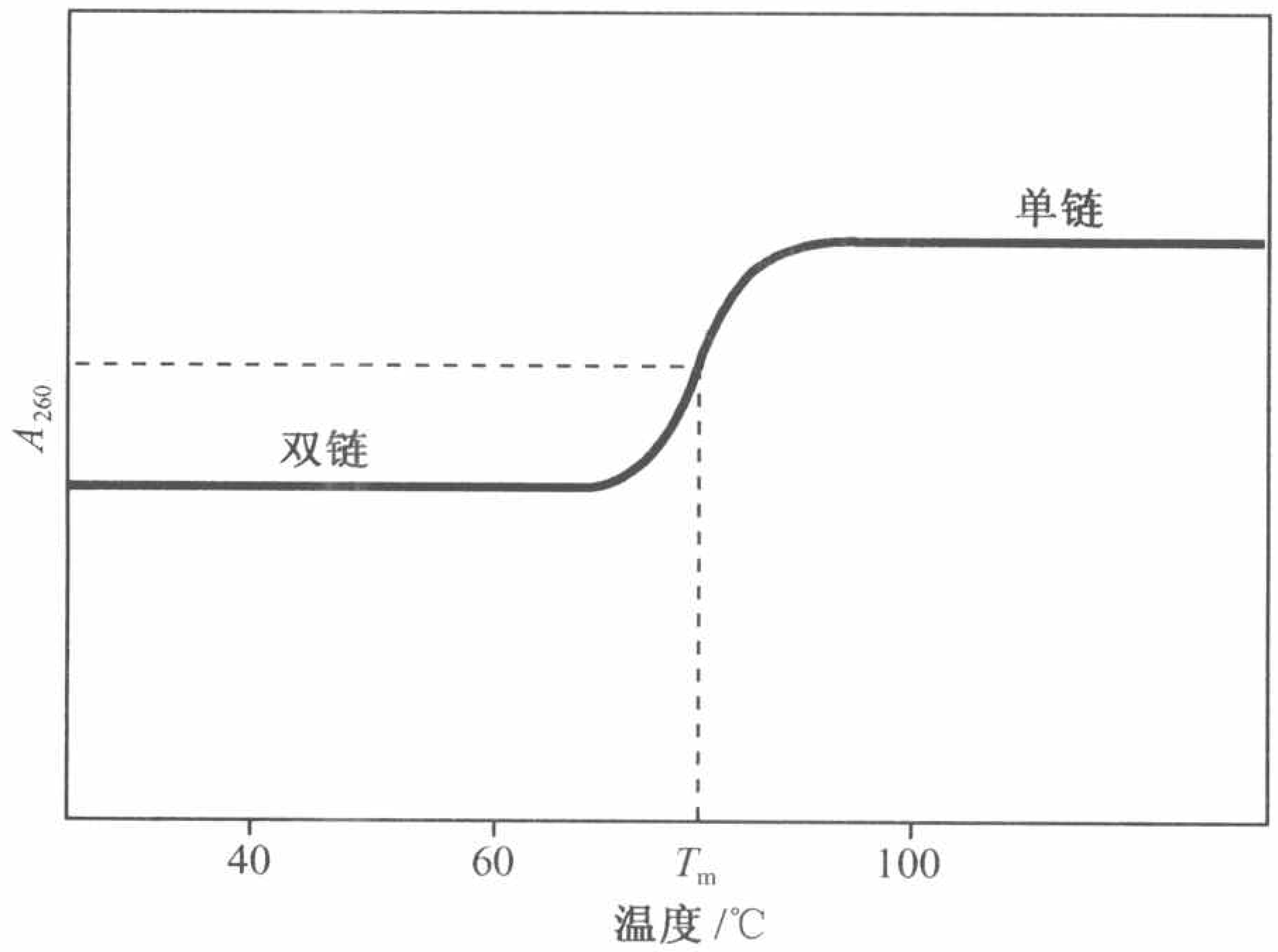


图 6-15 DNA 变性曲线。

T_m 值是 DNA 的特征常数，极大地取决于 DNA 中 G+C 的百分含量和溶液中的离子强度。DNA 中 G+C 的百分含量越高（即 A+T 的百分含量越低），DNA 的 T_m 值越高（图 6-16）。DNA 溶液的盐浓度越高，DNA 的 T_m 值也越高。这是为什么呢？与 A∶T碱基对相比，G∶C 碱基对对 DNA 的稳定性具有更重要的作用：首先，G∶C 碱基对之间有 3 个氢键，而 A∶T 碱基对之间只有 2 个氢键；更为重要的是，G∶C 碱基对与相邻碱基对之间的堆积力比 A∶T 碱基对与其相邻碱基对之间的堆积力更大。离子强度的效应则反映出双螺旋的另一基本特性。两条 DNA 链的骨架包含了带负电的磷酸基团。如果这些负电荷没有被中和，双链之间这种静电斥力将驱使两条链分开。在高离子强度下，负电荷可以被阳离子所中和，促进了双螺旋结构的稳定；相反，在低离子强度下，未被中和的负电荷将会降低双螺旋的稳定性。

一些 DNA 为环状分子

早期研究认为所有的 DNA 分子都是线形的，具有两个游离末端。确实，每个真核细胞的染色体都有一个极长的单独的 DNA 分子，但是现在我们知道某些 DNA 是环状的。例如，小猴 DNA 病毒 SV40 染色体就是环状双螺旋 DNA 分子，长 5000bp。并且大多数细菌染色体也是环状的。*E. coli* 的环状染色体长约 5000kb。此外，许多细菌都有染色体外自主复制的遗传因子——质粒（plasmid），通常为环状 DNA 分子。

有趣的是，一些 DNA 分子有时为线形有时为环状。最著名的例子是 λ 噬菌体，一种寄生在 *E. coli* 中的 DNA 病毒。在病毒粒子中 λ 噬菌体的基因组是线形双链分子。但当 λ 噬菌体感染 *E. coli* 细胞时，λ 噬菌体的基因组被注入 *E. coli* 细胞，DNA 分子随之

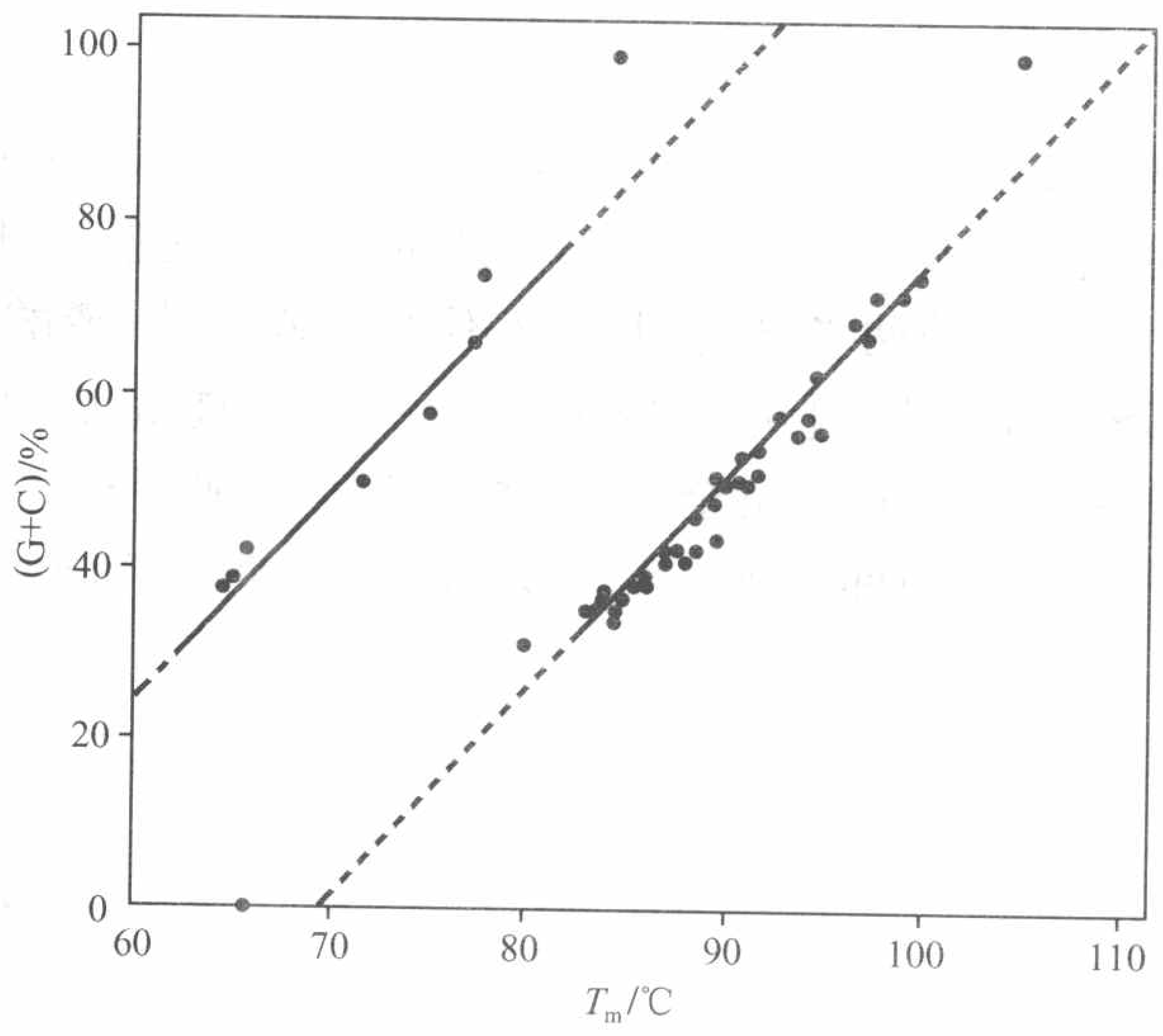


图 6-16 DNA 的变性取决于 G+C 含量和盐浓度。G+C 的百分含量越高，DNA 的变性温度（T_m）越高。不同来源的 DNA 分别溶解在低盐浓度（红线表示）和高盐浓度（绿线表示）的溶液中，pH 7.0。各点代表 DNA 的变性温度，以 G+C 的百分含量为纵坐标作图。（来源：Marmur J. and Doty P. 1962 Journal of Molecular Biology 5：120. © Elsevier）

发生环化。这是由于 λ 噬菌体的双链 DNA 末端具有互补的单链凸出部分，即“黏性末端”，可通过碱基配对连接而成为环状。

DNA 拓扑学

DNA 是一种柔性结构，它精确的分子参数与两方面相关：①周围的离子环境；②和它形成 DNA-蛋白质复合体的 DNA 结合蛋白的性质。因为线性 DNA 分子的末端是游离的，所以它们可以自由旋转以适应双螺旋两条链彼此缠绕次数的变化。但是如果 DNA 分子的两个末端共价连接形成环状分子，并且两条链的糖-磷酸骨架没有间断，DNA 分子的两条链彼此缠绕的次数就不能改变。这样的共价闭环 DNA 的拓扑结构是受限的。甚至由于长度极长，染色质结构以及与其他细胞成分的相互作用等原因，真核细胞染色体的线形 DNA 分子的拓扑结构也是受限的（第 7 章）。尽管有这些限制，DNA 仍然在细胞中参与了许多动力学过程。例如，双螺旋的相互缠绕的两条链必须迅速分开以完成 DNA 复制并转录为 RNA。因此，理解 DNA 拓扑学以及细胞如何在 DNA 复制、转录及其他的染色体活动中适应并利用拓扑结构的限制，在分子生物学中具有极其重要的意义。

连环数是共价闭环 DNA 的固有拓扑特性

考虑一下共价闭环 DNA（covalently closed，circular DNA，简称 cccDNA）的拓扑特性。因为多核苷酸链中无断裂，所以不破坏共价键就不能使 cccDNA 的两条链彼此分离。如果想使环状 DNA 的两条链分开而又不永久破坏糖-磷酸骨架上的任何共价键，我们必须使一条链反复穿过另一条链（有一种酶能完成这个过程）。连环数（linking number）是指要使两条链完全分开时一条链必须穿过另一条链的次数（图 6-17）。不论 DNA 分子如何扭曲，连环数是共价闭环 DNA 的固有拓扑特性，而且一定是整数。

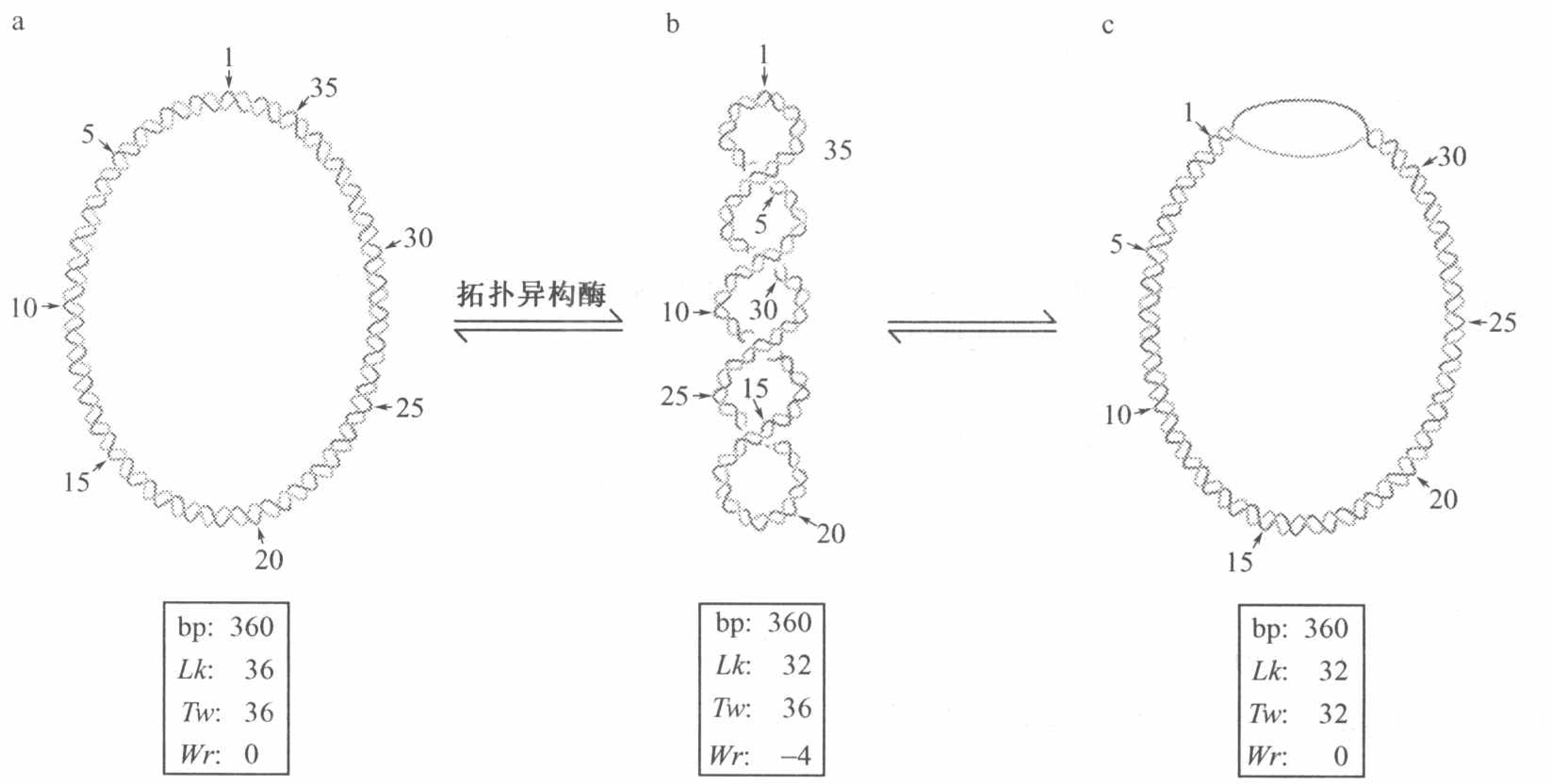


图 6-17　共价闭环 DNA（cccDNA）**的拓扑结构。**图示为 DNA 松弛态（a）到负超螺旋结构（b）的转变，超螺旋结构的张力可通过环状 DNA 形成超螺旋（b）或碱基对氢键的局部断裂（c）以消除。（据 Dr. M. Gellert 许可修改。）（来源：Modified from Kornberg A. and Baker T. A. 1992. DNA replication，11-21，p. 32. © W. H. Freeman.）

连环数由扭转数和缠绕数共同决定

连环数是扭转数（twist number）和缠绕数（writhe number）这两个几何数的总和（见交互动画 6-1）。先来看扭转数，扭转数仅仅是一条链旋绕另一条链的螺旋数，即一条链完全缠绕另一条链的次数。设想一个 cccDNA 分子平躺在一个平面上，在这个平面构象中，连环数就等于扭转数。确实，扭转数可以由统计两条链彼此缠绕的次数而加以确认（图 6-17a）。右手螺旋的螺旋交叉数（扭转数）被定义为正数，因此 DNA 的连环数是正值。

但是 cccDNA 通常不是处在一个平面上，而是扭转的，以至于在三维空间里双螺旋的长轴经常重复地自我交叉（图 6-17b），这称为缠绕数。为使这种外力造成的扭曲更加形象化，可以想像一下过度扭曲的电话线。

缠绕可能有两种形式：一种形式称互卷（interwound）或相缠螺旋（plectonemic writhe），如图 6-17b 和图 6-18a 描述的，长轴本身自我交叉；另一种形式为环形（toroid）或螺线管状（spiral），像通常 DNA 缠绕蛋白质时一样，其长轴以螺线管形式绕成圆柱状（图 6-18b ）。缠绕数（Wr），是 cccDNA 中相缠螺旋形式缠绕数和（或）螺线管形式缠绕数的总和。例如，图 6-17b 所示分子的缠绕数为 4。

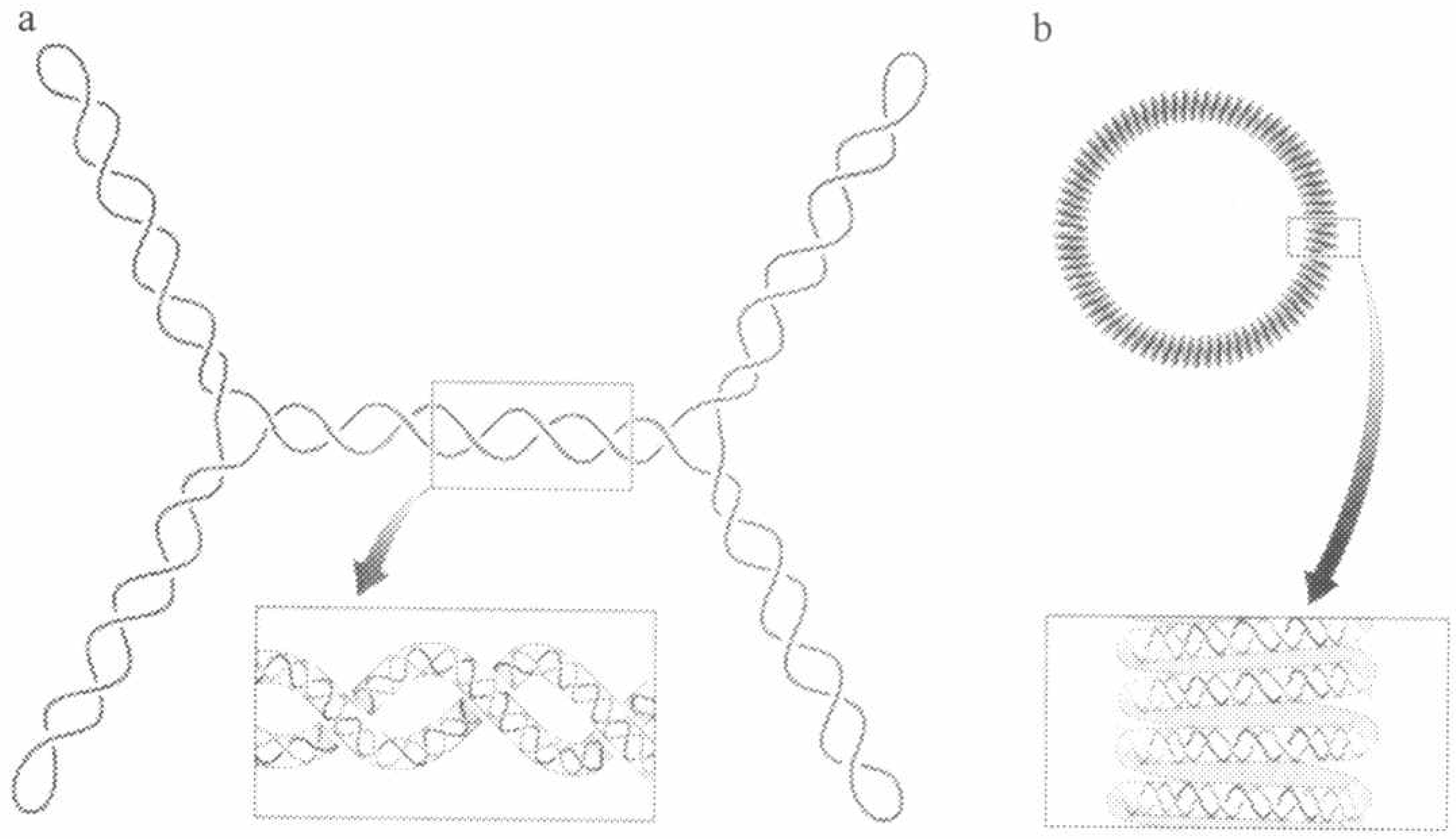


图 6-18 超螺旋 DNA 的两种缠绕方式。图示为同等长度 cccDNA 的相缠螺旋（a）和环形螺线管（b）两种缠绕方式。a. 双螺旋 DNA 自身发生扭曲形成互卷或称相缠螺旋，如图中所示的支链分子。b. 环形或螺线管缠绕被描述为圆柱形盘绕。（来源：据 Kornberg A. and Baker T. A. 1992. DNA replication, 11-21, p. 33. © 1992 by W. H. Freeman and Company. 许可修改，经 Dr. Nicholas Cozzarelli 许可）

相缠螺旋形式和螺线管形式是 cccDNA 拓扑等价的几何特性，容易发生相互转化，扭转和缠绕也是可以相互转化的。一个 cccDNA 分子容易发生扭转变形，在不破坏任何共价键的条件下，部分扭转转变为缠绕或者部分缠绕转变为扭转。唯一不变的是扭转数（Tw）和缠绕数（Wr）的总和必须与连环数（Lk）相等，用等式表示如下：

$$Lk = Tw + Wr$$

Lk° 是生理条件下完全松弛态 cccDNA 的连环数

设想 cccDNA 未形成超螺旋（supercoiling），即松弛态（relaxed），则它的扭转数就与生理条件下 DNA 溶液中的 B-DNA 相同（每螺周 10.5 个碱基对）。生理条件下 cccDNA 的连环数（Lk）表示为 Lk°。这样一个分子的 Lk° 就是碱基对数除以 10.5 所得的值。假如一个 cccDNA 长度为 10 500bp，则 $Lk=+1000$（由于 DNA 是右手螺旋，此值为正数）。为了说明这一点，我们不妨做个实验，将长度为 10 500bp 的 cccDNA 的一条链拉展为平面环状，则另一条链与这条链交叉 1000 次。

如果 cccDNA 不处于松弛态我们如何消除它的超螺旋呢？一个方法是用 DNase Ⅰ温和处理 DNA，目的是平均每个 DNA 分子打断一个磷酸二酯键（或一小部分键）。一旦 DNA 以这种方式产生“缺口”就不再受拓扑抑制，它的链可以自由旋转，使缠绕消失（图

6-19)。如果缺口被修复，产生的 cccDNA 分子呈松弛态，Lk 平均值与 Lk° 相等（由于修复缺口时旋转程度波动，所产生的 cccDNA 的 Lk 有的略大于 Lk°，有的略小于 Lk°。因此这种解旋的方法将会产生一系列限定在很窄范围内的 cccDNA，其 Lk 平均值与 Lk° 相等)。

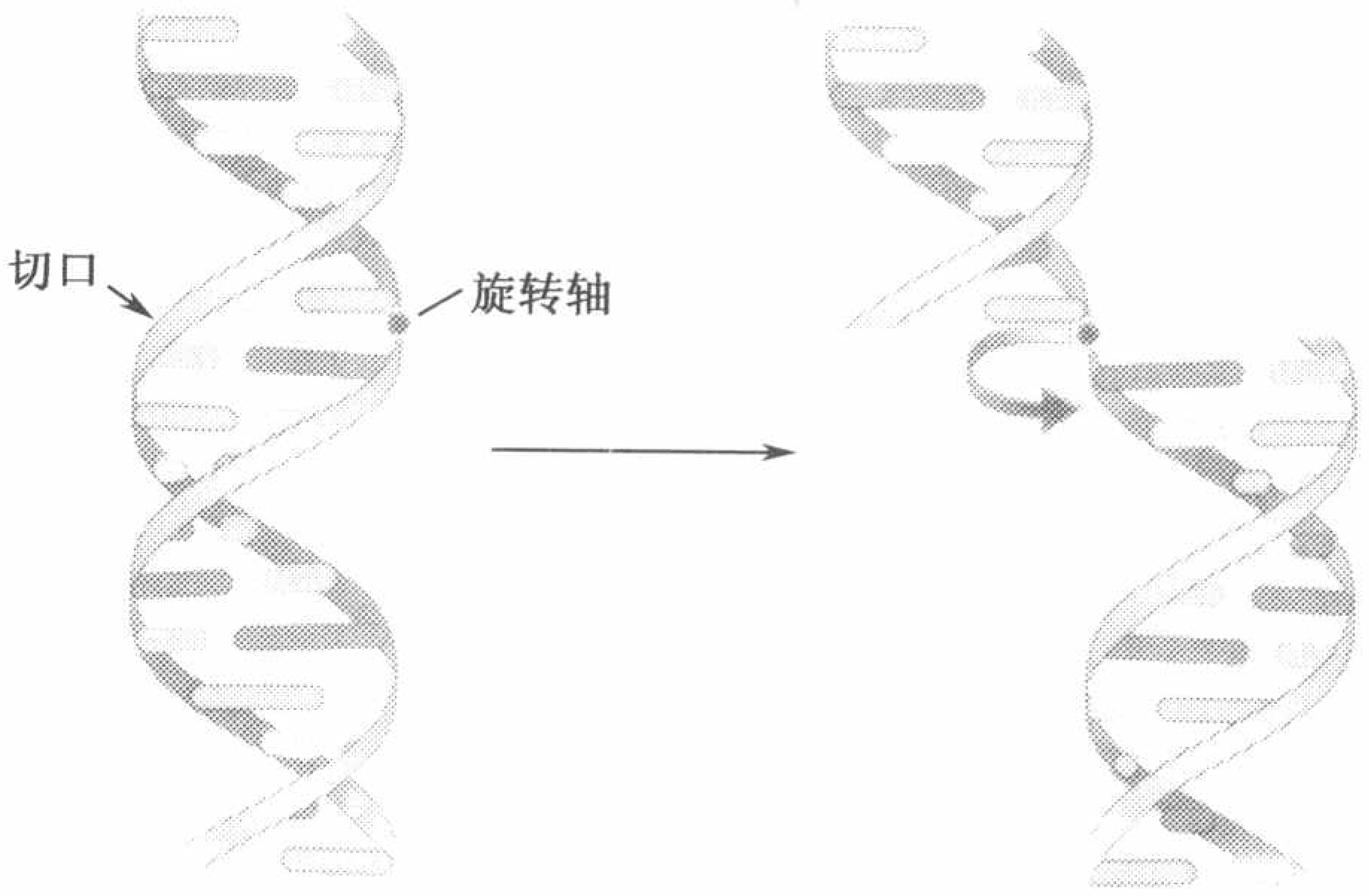


图 6-19 DNase Ⅰ作用可使 DNA 解旋。

细胞中 DNA 呈负超螺旋

用 Lk 与 Lk° 的差值表示超螺旋的程度，称为连环差（linking difference）：

$$\Delta Lk = Lk - Lk^{\circ}$$

如果 cccDNA 的 ΔLk 绝对值远大于 0，DNA 就会因为过度扭曲产生超螺旋。$Lk < Lk^{\circ}$，即 $\Delta Lk < 0$，DNA 产生“负超螺旋”；$Lk > Lk^{\circ}$，即 $\Delta Lk > 0$，DNA 产生“正超螺旋”。举例来说，图 6-17a 所示分子为松弛态 $Lk^{\circ} = 36$，图 6-17b 所示分子为负超螺旋，$Lk = 32$，后者与前者相差 4，所以 $\Delta Lk = -4$。

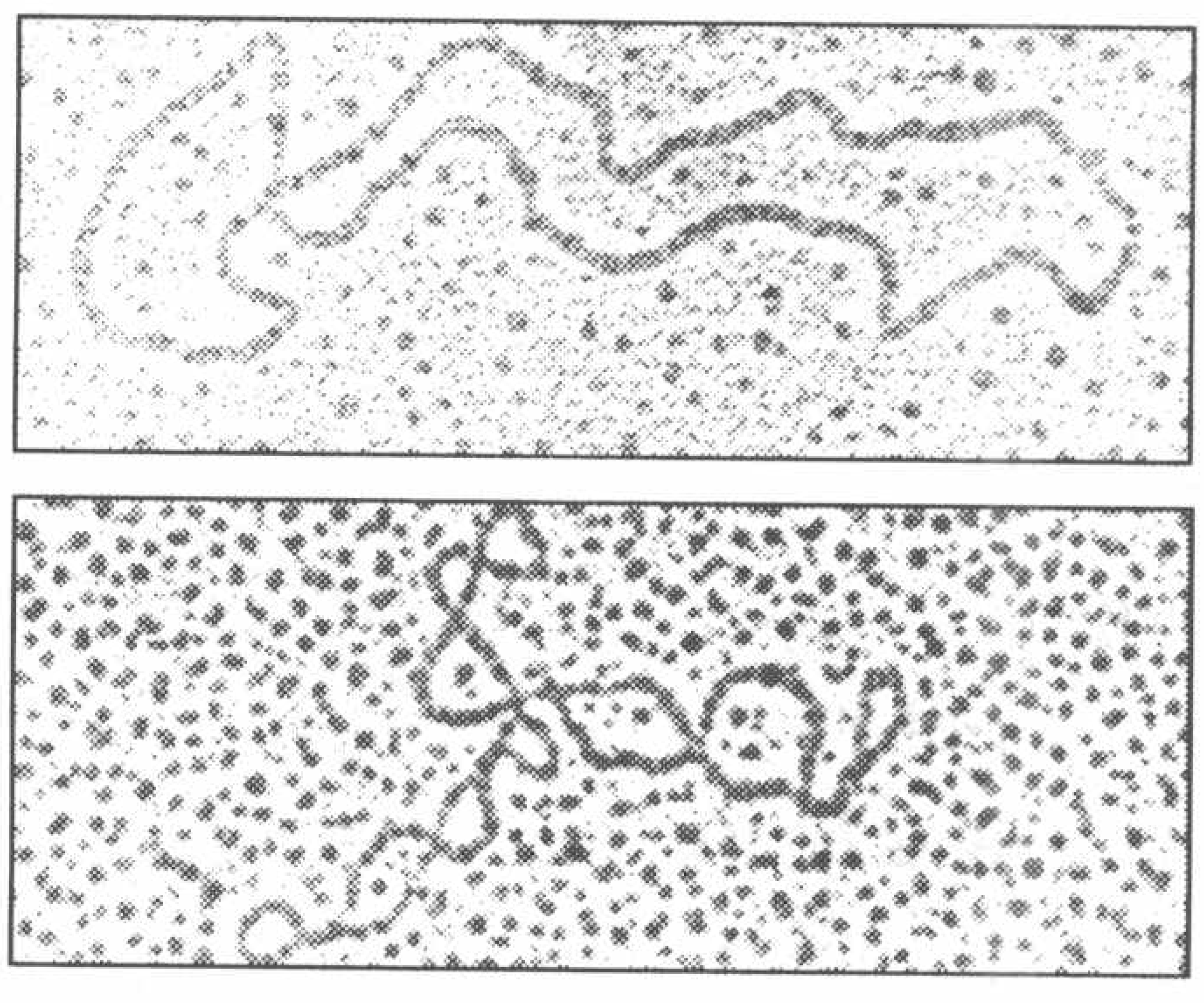


图 6-20 超螺旋 DNA 的电子显微照片。上图为噬菌体 PM2 的松弛态（未发生超螺旋）DNA 的电子显微照片；下图为噬菌体超螺旋形式的电子显微照片。（来源：Electron micrographs courtesy of Wang J. C. 1982，Scientific American 247：97.）

因为 ΔLk 与 Lk° 取决于 DNA 分子的长度，采用一个超螺旋的标准测量方式会更方便。这就是超螺旋密度（superhelical density），用 σ 表示，定义如下：

$$\sigma = \Delta Lk / Lk^{\circ}$$

从细菌和真核细胞中提取的环状 DNA 分子通常是负超螺旋，σ 值约为 −0.06。图 6-20 所示电子显微照片将噬菌体 DNA 松弛态结构与超螺旋结构进行了比较。

超螺旋的生物学意义是什么呢？负超螺旋含有自由能，可以为打开双链提供能量，使 DNA 复制和转录这样的解离双链的过程得以顺利完成。因为 $Lk = Tw + Wr$，负超螺旋可以让双螺旋扭转数减少（比较图 6-17a 和图

6-17b)，负超螺旋区域因而有局部解旋倾向。因此，与松弛态 DNA 相比，负超螺旋 DNA 更易于解链。

唯一含有正超螺旋 DNA 的生物体是某些生活在极端高温环境（如温泉）中的嗜热微生物。在这种情况下，我们认为正超螺旋提供能量，阻止 DNA 在高温中发生变性。正超螺旋多转几圈（DNA 过度扭曲），则嗜热微生物 DNA 双链打开就比一般生物呈负超螺旋的 DNA 需要更多能量。

真核细胞中核小体引入了负超螺旋

下一章中我们将提到真核细胞细胞核中的 DNA 被包装成小颗粒，称为核小体 (nucleosome)。在核小体中，DNA 双螺旋以螺线管方式围绕蛋白质核心缠绕两圈。重要的是，它以左手螺旋方式进行盘绕（伸出左手，想像 DNA 如何盘绕在核小体上，见第 7 章，图 7-18)。实际上，左手螺线管方式就等同于负超螺旋。因此，DNA 包装成核小体引入了负超螺旋密度。

拓扑异构酶可使超螺旋 DNA 解旋

连环数是 DNA 的固有特性，受制于 DNA 的拓扑结构。只有破坏糖-磷酸骨架才能使它发生改变。拓扑异构酶（topoisomerase）可以催化 DNA 产生瞬时单链或双链的断裂而改变连环数。

拓扑异构酶有两种基本类型。拓扑异构酶Ⅱ通过两步改变 DNA 连环数。它们在 DNA 上产生一瞬时的双链缺口，并在缺口闭合以前使一小段未被切割的双螺旋 DNA 穿过这一缺口，图 6-21 是此反应的大概过程示意。拓扑异构酶Ⅱ依靠 ATP 水解提供的能量来催化这一反应。相反，拓扑异构酶Ⅰ一步就可以改变 DNA 的连环数，其作用是使 DNA 暂时产生单链切口，让未被切割的另一条单链在切口闭合之前穿过这一切口（图 6-22)。与拓扑异构酶Ⅱ相比，拓扑异构酶Ⅰ不需要 ATP。下面将阐述拓扑异构酶如何以受控的、协调的方式松弛 DNA 以及促进其他相关反应。

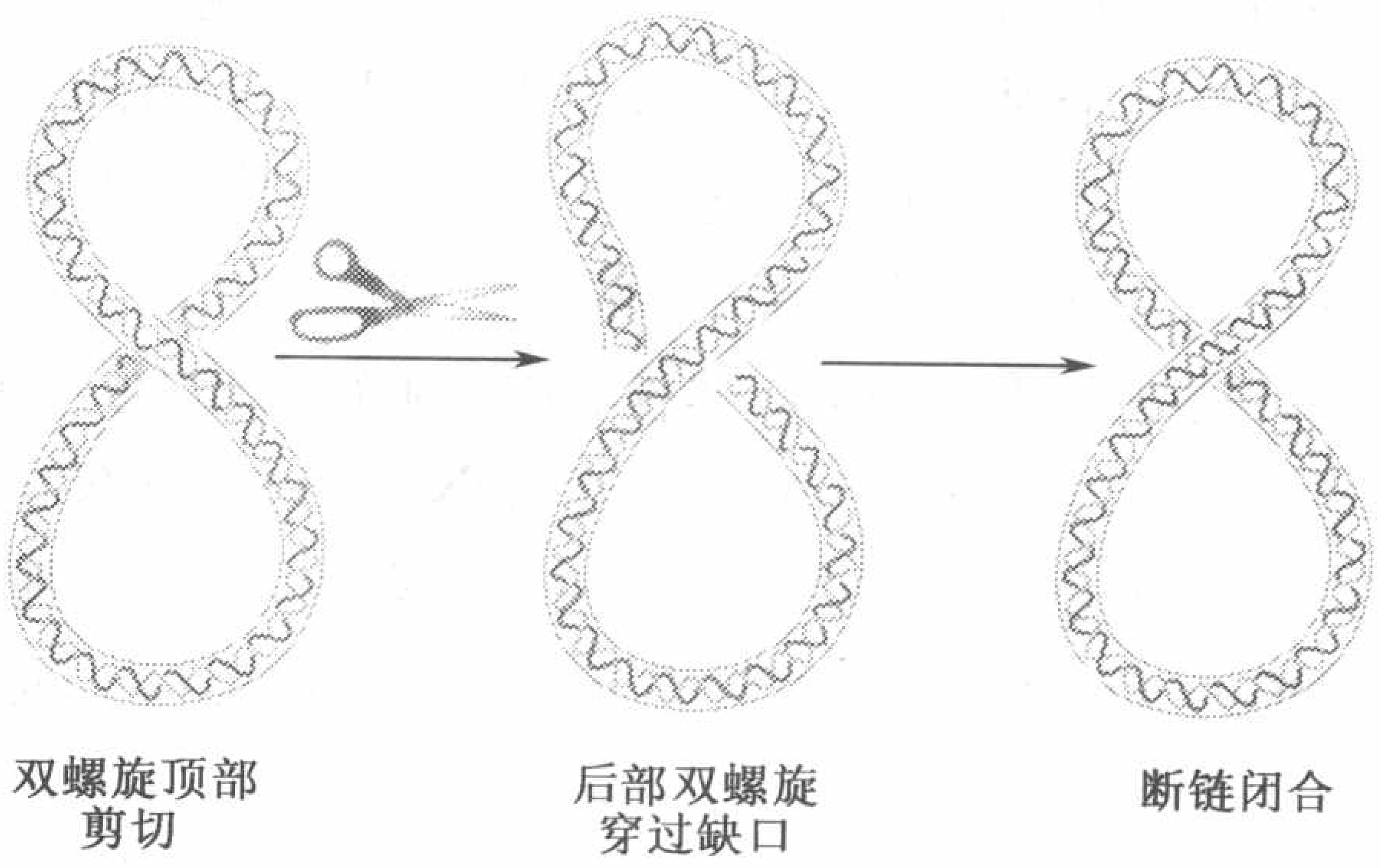


图 6-21 DNA 在拓扑异构酶Ⅱ作用下改变连环数的过程。拓扑异构酶Ⅱ结合于 DNA，产生双链断裂，未断裂的 DNA 穿过缺口后，再将断链闭合。

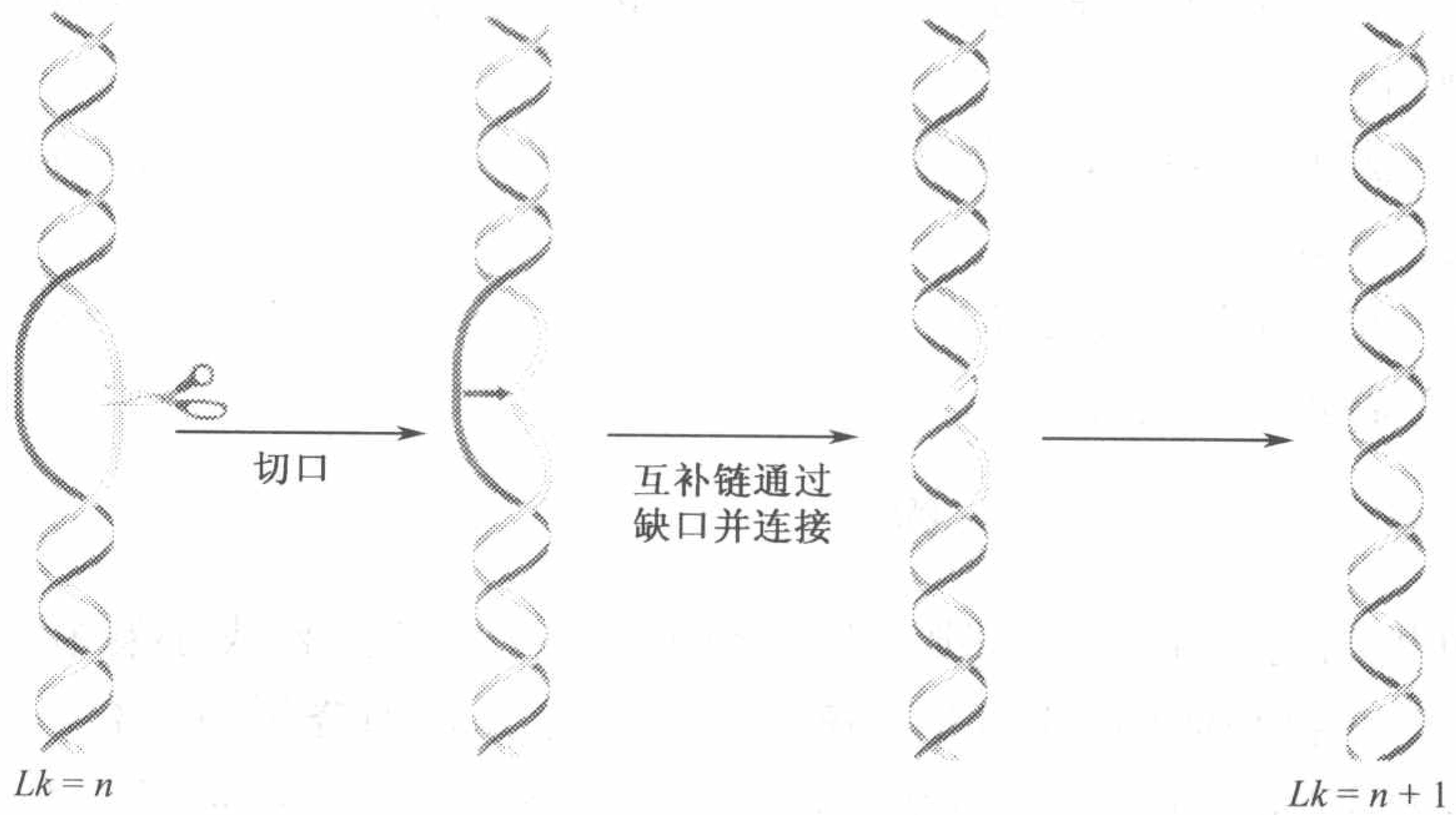


图 6-22 拓扑异构酶Ⅰ的作用机制图。拓扑异构酶切开 DNA 双链中的一条，没有被切的互补链通过缺口穿过来，然后连上缺口，该过程使连环数加 1。

原核生物中存在引导 DNA 超螺旋形成的特殊拓扑异构酶

原核和真核生物都存在可以从 DNA 中去除超螺旋的拓扑异构酶Ⅰ和Ⅱ。另外，原核生物还拥有一种特殊的拓扑异构酶Ⅱ，通称为促旋酶，它引入而不是去除负超螺旋。DNA 促旋酶使原核生物染色体产生负超螺旋。这种负超螺旋促进 DNA 双螺旋的解旋，而这种解旋又会激发包括启动 DNA 转录和复制在内的许多反应。

拓扑异构酶也可以解链和松弛 DNA 分子

拓扑异构酶不仅能松弛双螺旋 DNA，而且能促进一些与维持细胞内 DNA 的适当结构有关的重要反应的发生。这些酶利用松弛双螺旋 DNA 相同的瞬时 DNA 断裂以及链通过反应机制来实现这些反应。

拓扑异构酶既能链接（catenate）也能解开（decatenate）环状 DNA 分子。环状 DNA 分子如果像一条链的两个环一样连接在一起，就称为相互“链接”（图 6-23a）。在拓扑异构酶这两个活性中，它解开环状 DNA 的能力显然更富有生物学意义。如同我们将在第 8 章看到的，一轮 DNA 复制完成后通常产生链接在一起的 DNA 分子（图 8-34 所示）。拓扑异构酶在解开这些 DNA，以便使它们在细胞分离时分配到两个子细胞中去的过程中起关键作用。两个共价闭合的环状 DNA 分子的解开要求一个 DNA 分子的两条链穿过另一个 DNA 分子的双链缺口。这些反应有赖于拓扑异构酶Ⅱ。由于解链的必要性，拓扑异构酶Ⅱ是细胞必需的蛋白质。然而，如果两个相互链接的 DNA 分子中至少有一个携带缺口，那么拓扑异构酶Ⅰ也可以解开这两个分子（图 6-23b）。

虽然在考虑拓扑问题的时候，我们关注的总是环状 DNA 分子，真核生物的长线性染色体也存在拓扑结构。例如，在每一轮 DNA 复制中，两条双链的子 DNA 分子通常缠绕在一起（图 6-23c）。这些相互缠绕的位点就像链接在一起的 DNA 分子之间的连接一样，阻碍了姐妹染色体在有丝分裂过程中的分离。因此，通常由拓扑异构酶Ⅱ催化的

DNA 解缠绕，也是 DNA 的成功复制和真核细胞的分裂所必需的。

有时，一个 DNA 分子会出现打结的现象（图 6-23d）。例如，第 11 章中我们将详细讨论的一些位点特异性重组反应会引起 DNA 的打结现象。在此，拓扑异构酶Ⅱ也可以解开这种双链 DNA 中的结。如果 DNA 分子中含有缺口，那么拓扑异构酶Ⅰ也可以完成解链工作。

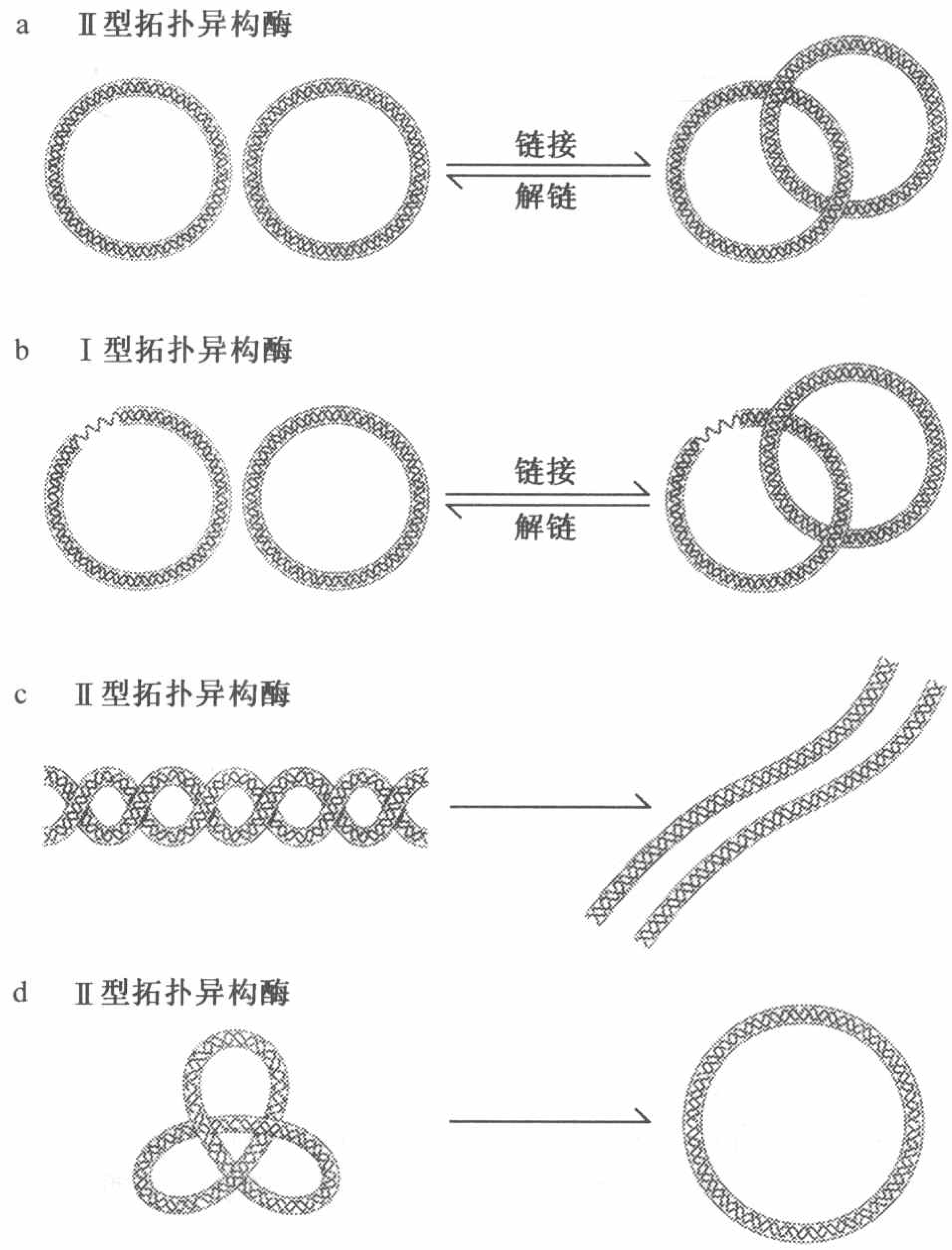


图 6-23 拓扑异构酶对 DNA 的解环、解旋和解结作用。 a. 拓扑异构酶Ⅱ能够在共价闭合环状 DNA 分子上产生一个双链缺口并使另一个 DNA 分子穿过缺口形成套在一起的双环链接或解链。b. 拓扑异构酶Ⅰ只有在 DNA 分子有缺口或断点的情况下才能使 DNA 分子链接或解链，因为这种酶一次只能切断一条 DNA 链。c. 真核生物染色体复制过程中所产生的相互缠绕的线性 DNA 长链能够被拓扑异构酶解开。d. DNA 的结能够被拓扑异构酶解开。

拓扑异构酶通过蛋白质-DNA 的共价连接裂解 DNA 链或使它们重新连接

拓扑异构酶为了发挥功能，必须能够切断一条或两条 DNA 链，然后再将切断的 DNA 链重新连接起来。通过共价中间物的机制，拓扑异构酶可以不依赖任何其他蛋白质或高能辅助因子（如 ATP；可以在下文看到）。当拓扑异构酶活性部位的酪氨酸残基

攻击靶 DNA 骨架上的磷酸二酯键时，可使 DNA 断裂产生一缺口（图 6-24），而拓扑异构酶则通过磷酸-酪氨酸连接与其中一个断端共价相连。DNA 的另一端则带有一个游离的羟基，可在下文中看到这个末端被酶如何紧紧地抓住。

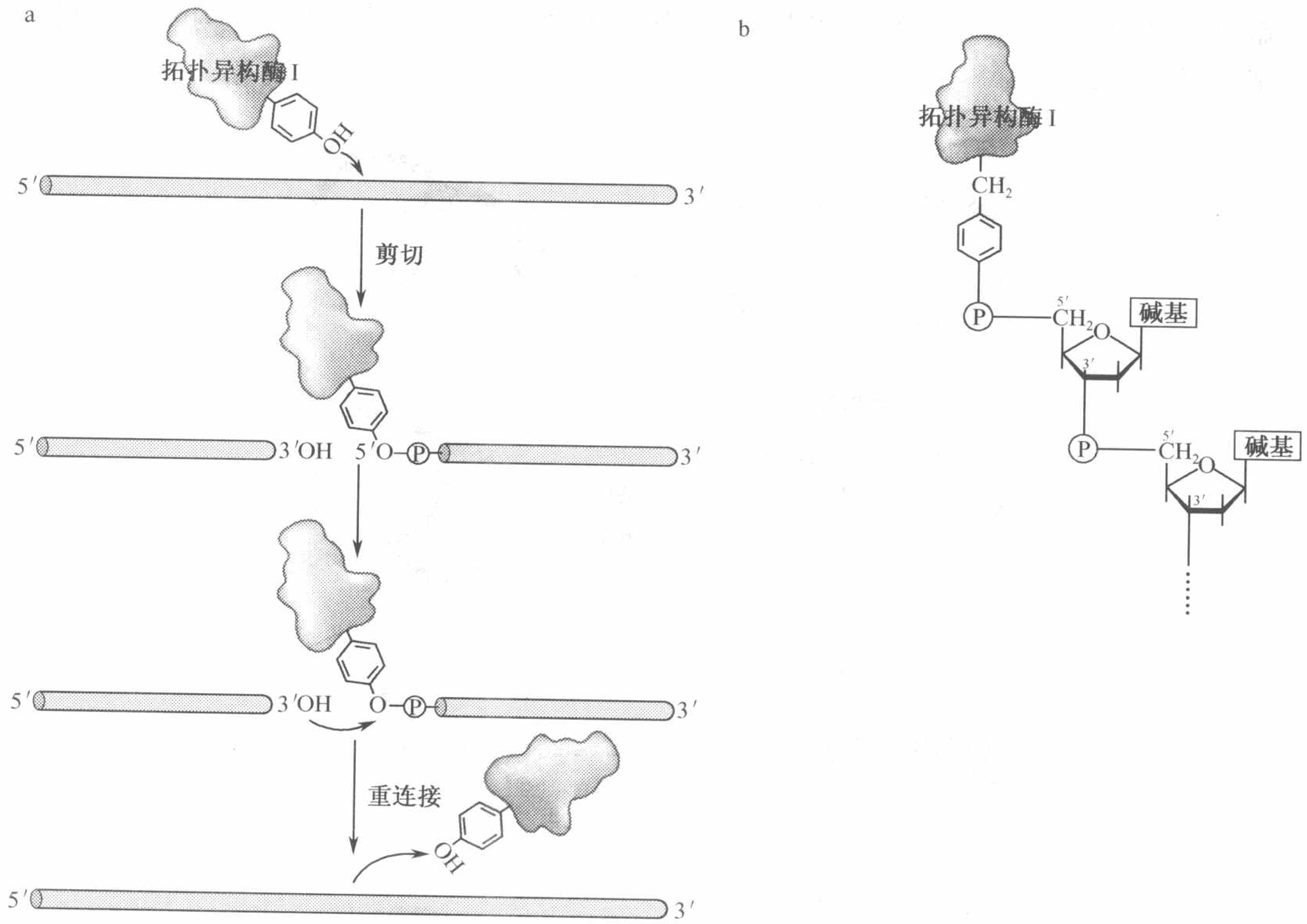


图 6-24 拓扑异构酶通过酪氨酸-DNA 共价中间体切开 DNA。 a. 断开和重连接反应。为简便起见，这里仅显示单条 DNA 链，图 6-25 可见更真实的图片。拓扑异构酶 Ⅱ 使用同样的机制，但需要两个酶的亚基，每个亚基打开一条 DNA 链。拓扑异构酶有时打开 DNA 5′端，有时打开 3′端。b. 磷酸-酪氨酸共价中间体的详细结构。

这种磷酸-酪氨酸连接保存了被裂解的磷酸二酯键的能量。因此，只要逆转原来的反应，DNA 就可以被重新闭合：来自 DNA 断端的游离羟基攻击磷酸-酪氨酸键，最终重新形成磷酸二酯键。这个反应重新连接 DNA 链，释放拓扑异构酶，拓扑异构酶可以接着催化另一轮反应。虽然像上面所说的，拓扑异构酶 Ⅱ 需要 ATP 水解释放的能量，这些能量主要用于拓扑异构酶- DNA 复合物的分子构象的改变，而并非使 DNA 链断裂或重新连接。

拓扑异构酶构成“酶桥”让 DNA 片段往来穿梭

在 DNA 断裂和重新连接的过程中，拓扑异构酶促使 DNA 的另一条片段穿过断点。因此，拓扑异构酶功能的行使要求 DNA 断裂，DNA 链穿过和 DNA 重新连接都以高度协调的方式发生。从几个不同拓扑异构酶的结构可以看出这些反应是怎么发生的。这里我们将解释拓扑异构酶 Ⅰ 松弛 DNA 的一个模型。

起始解链循环时，拓扑异构酶与 DNA 片段形成紧密的复合物，复合物中双链变性

(图 6-25a)。高度的负超螺旋 DNA 链倾向于熔解变性（如前面所述），这使得这种 DNA 成为松弛反应的良好底物。一条 DNA 链结合于酶的裂隙中，这使它靠近酪氨酸活性部位这条链断裂，产生 DNA-酪氨酸中间产物（图 6-25b）。这一步反应成功的关键在于新断裂的 DNA 的另一末端也被拓扑异构酶紧密结合。断裂后，拓扑异构酶经过一次大的构象变化，在断裂的链中打开一个缺口，并通过自身成桥状连接这个缺口。第二条链（没有断裂的那条链）此时穿过那个缺口，结合于蛋白质内部的“圆环状”孔中的 DNA 结合位点（图 6-25c）。链穿过以后，拓扑异构酶-DNA 复合物中发生第二次构象变化，通过攻击磷酸-酪氨酸键的羟基（如前所述）使 DNA 的断端重新连接起来（图 6-25d）。重新连接以后，拓扑异构酶必须再经过一次构象变化，释放 DNA（图 6-25e）。这一过程的 DNA 产物与最初的 DNA 分子基本相同，只是连环数增加了一个。

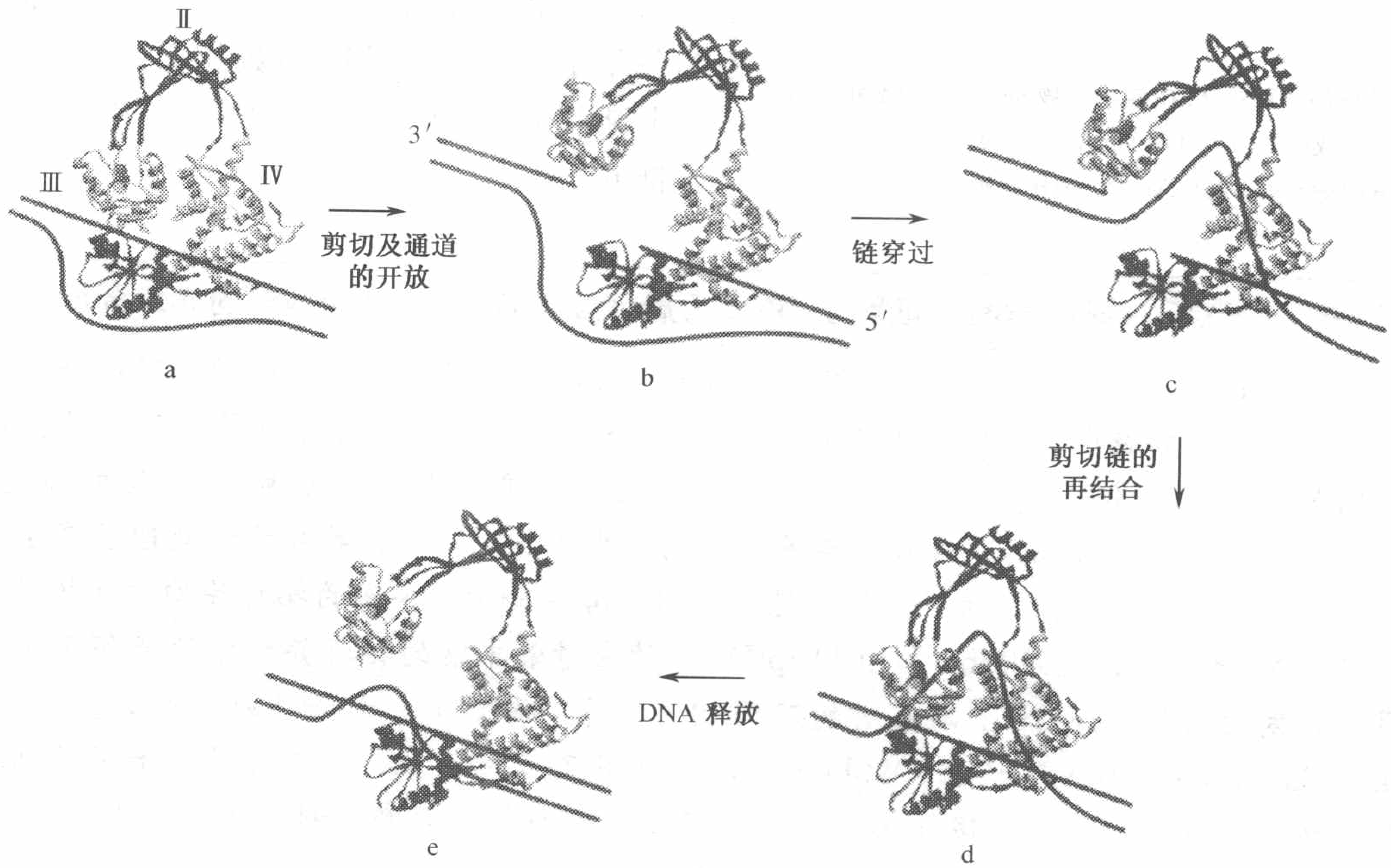


图 6-25 拓扑异构酶Ⅰ催化反应模型。图示为推测的质粒 DNA 的负超螺旋一系列的解旋过程。DNA 两条链如黑灰线所示（不按比例），蛋白质的四个结构域按区域标记（a）。结构域Ⅰ用红色，Ⅱ用蓝色，Ⅲ用绿色，Ⅳ用橙色标记。（来源：Champoux J. 2001. DNA topoisomerases. Annual Review of Biochemistry 70：369-413. © Annual Reviews.）

上述反应机制中拓扑异构酶在链穿行过程中起到了“蛋白桥”的作用，这种机制也适用于拓扑异构酶Ⅱ。然而拓扑异构酶Ⅱ是二聚体（某些情况下是四聚体）。拓扑异构酶Ⅱ作用机制的基本特点是 DNA 双链的断裂需要拓扑异构酶两个亚基及其活性位点中的酪氨酸残基的参与。

DNA 拓扑异构体可被电泳分离

长度相同而连环数不同的共价闭环 DNA 分子叫做 DNA 拓扑异构体（DNA topoi-

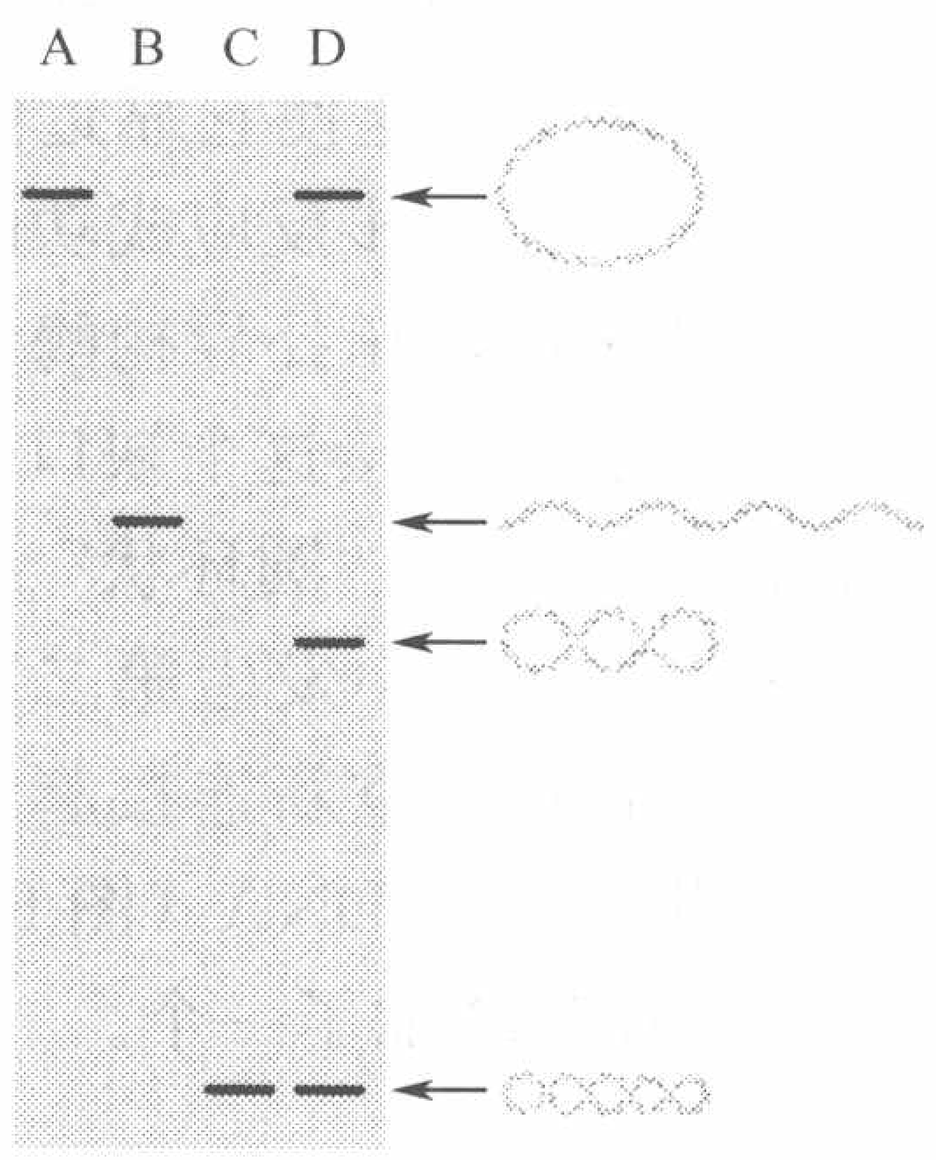


图 6-26 DNA 拓扑异构产物的电泳分离图。A：松弛或有缺口的环状 DNA；B：线性 DNA；C：高度超螺旋的 cccDNA；D：拓扑异构体梯度标记。

somer)。虽然拓扑异构体的分子量相同，它们也可以通过琼脂糖凝胶电泳［见第 20 章的凝胶电泳（gel electrophoresis）的解释］彼此分开。拓扑异构体分离的基础是扭曲越大，cccDNA 的形状就越致密。在此，可以再次试想电话线的超螺旋如何使它变得更致密。DNA 越致密，就越容易在凝胶基质中迁移（图 6-26）。因此，一个完全松弛的 cccDNA 的迁移要比这一环状 DNA 链的高度超螺旋的拓扑结构体慢得多。图 6-27 所示是 DNA 拓扑结构体被凝胶电泳分离成一系列条带。相邻条带的 DNA 分子连环数相差一个。显然，电泳迁移率与 DNA 的拓扑结构密切相关（框 6-3）。

框 6-3 DNA 环的拓扑学特性证明了 DNA 每周螺旋含有 10.5 个碱基对的螺旋周期性

由于 DNA 的不同拓扑结构体能通过电泳分离，人们根据这一发现设计了一个简单的实验来证明溶液中 DNA 一周螺旋含有 10.5 个碱基对。假设用拓扑异构酶Ⅰ将大小分别为 3990bp、3995bp 和 4011bp 的三种 cccDNA 完全解旋。琼脂糖凝胶电泳时，3990bp 和 4011bp 的 cccDNA 迁移率基本相同。由于热波动，拓扑异构酶处理后产生了一系列拓扑异构体，但是为了方便起见，我们只考虑含量最丰富的拓扑异构体（相对于最松弛态 cccDNA)，3990bp 和 4011bp DNA 的含量最丰富的拓扑异构体的迁移率难以区分，因为它们之间 21 个碱基对的不同相对整个 DNA 环来说是可以忽略的。然而，虽然 3995bp 的 DNA 环只比 3990bp 的 DNA 环多了 5 个碱基，但其含量最丰富的拓扑异构体却比另外两种环迁移得稍快一些。那么，我们怎样解释这种现象呢？3990bp 和 4011bp 的 DNA 环处于最松弛的状态时，连环数正好等于 Lk^{o}，对于 3990bp 的 DNA 来说（大小按 10.5 个碱基对分开）Lk^{o} 等于 380，而对于 4011bp 的 DNA 来说 Lk^{o} 等于 382。因为 Lk 等于 Lk^{o}，这两者的连环数的差值（$\Delta Lk = Lk - Lk^{o}$）都为 0，DNA 不存在扭曲。由于连环数必须是整数，而 3995bp 的 DNA 环在最松弛的状态下可以有连环数为 380 或 381 的两种拓扑异构体。然而，3995bp 的 DNA 环的 Lk^{o} 是 380.5。这样，即使在它最松弛的状态下，3995bp 的共价闭合环状 DNA 也必须有半个单位的扭曲（连环数差值为 0.5)，因此它比 3990bp 和 4011bp 的 DNA 环迁移得快。换句话说，为了解释长度相差 21 个碱基对（螺旋的两圈）的两个 DNA 环迁移率相同，而长度相差只有 5 个碱基对（大约一半螺旋）的两个 DNA 环迁移率却不同，我们得出了“溶液中的 DNA 每周螺旋含有 10.5 个碱基”的结论。

溴乙锭离子可使 DNA 解旋

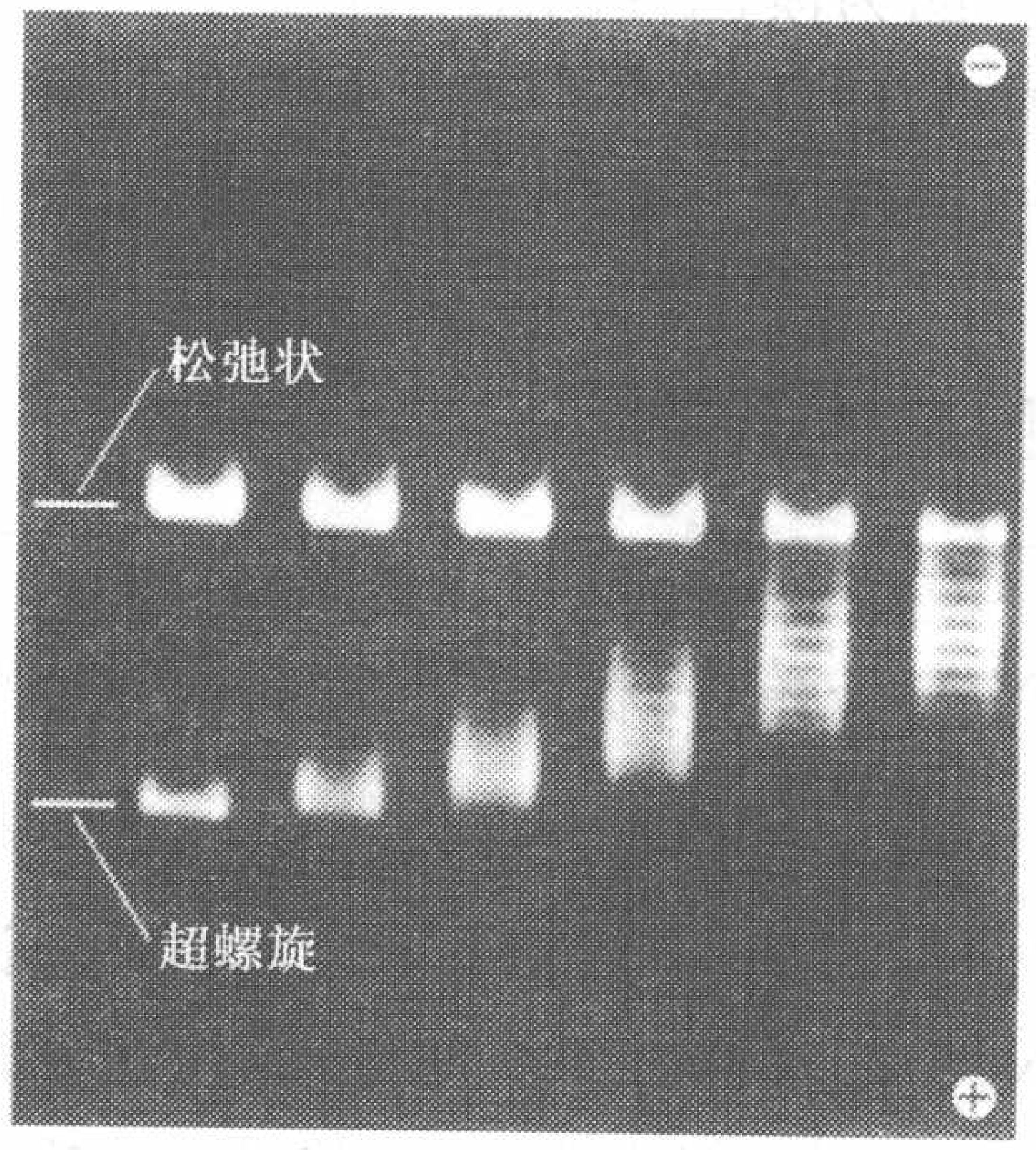


图 6-27 用凝胶电泳分离超螺旋和松弛的 DNA。用凝胶电泳分离超螺旋和松弛的 DNA。电泳的速度按超螺旋圈数的增加而增加。(来源：J. C. Wang 惠赠)

溴乙锭（ethidium）是一种既大又平而且多环的阳离子，它的平面形状使其能滑入或嵌入到堆积的 DNA 碱基对平面之间（图6-28）。因为在紫外光下溴乙锭会发荧光，而且嵌入 DNA 后荧光强度显著增强，因此，溴乙锭常用来作为染料检测 DNA。

当溴乙锭嵌入两个碱基对之间时，引起 DNA 解旋 26°，使得通常的碱基对之间的夹角由 36°变为 10°，换句话说，溴乙锭降低了 DNA 的扭曲程度。想像一下若 DNA 分子每个碱基对中都嵌有一个溴乙锭离子，那么每周螺旋不再是 10 个碱基对而是 36 个碱基对。当溴乙锭结合的是线性或缺口环状 DNA 分子时，溴乙锭只引起螺旋倾斜度的增加。但是当溴乙锭结合于共价闭合环状 DNA 时结果会怎么样呢？cccDNA 连环数并没有增加（并没有共价键断裂或形成），但是每分子溴乙锭嵌入则使每圈的扭曲度降低 26°。因为$Lk=Tw+Wr$，这样，Tw 的减少必须由 Wr 的增加来补偿。如果最初环状 DNA 为负超螺旋（细胞中分离出的环状 DNA 大多如此），加入溴乙锭将增加 Wr，也就是说，溴乙锭的加入可松弛 DNA。如果加入足量的溴乙锭，负超螺旋将变为 0；若加入更多的溴乙锭 Wr 将大于 0，此时 DNA 将变为正超螺旋。

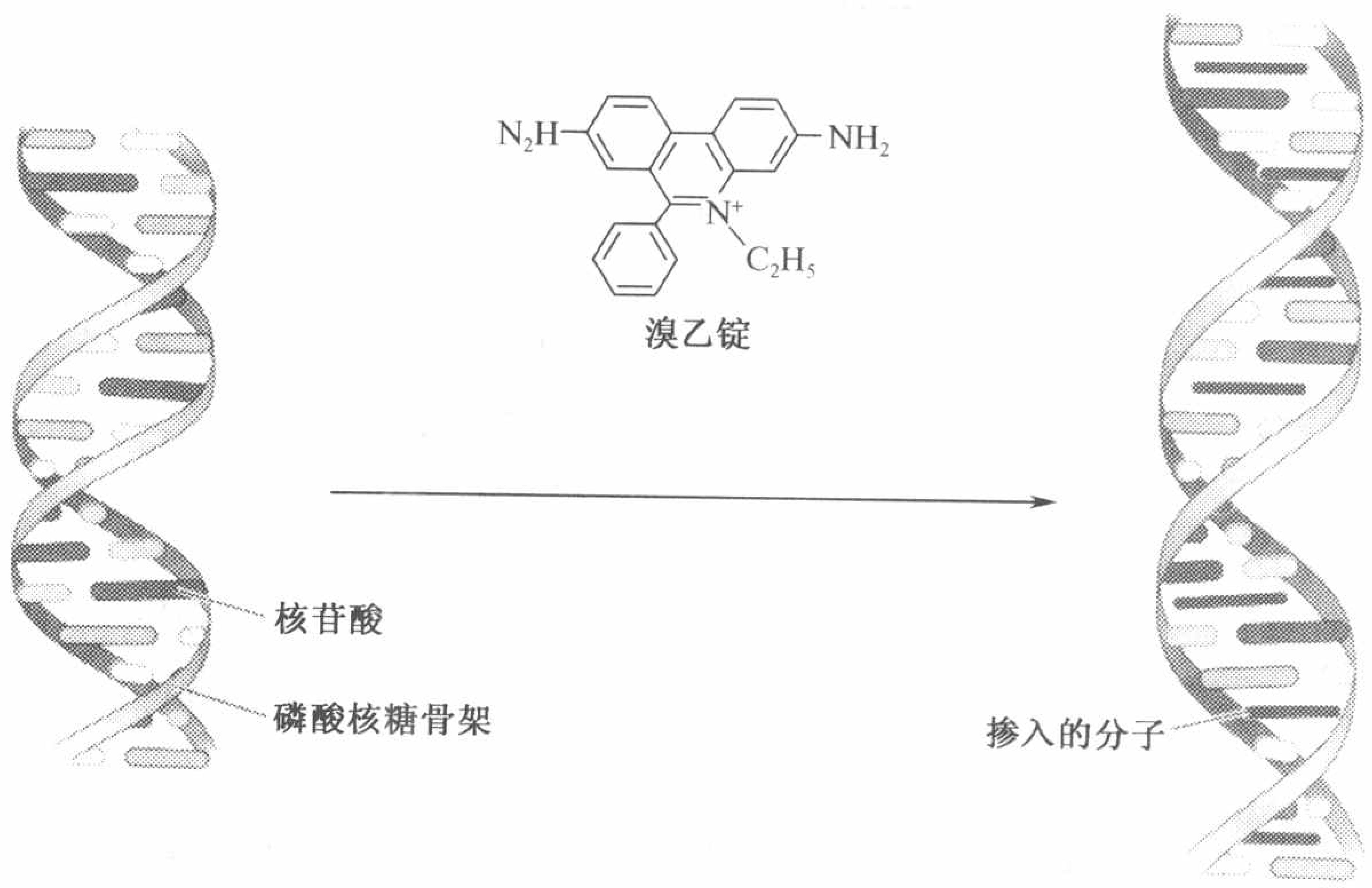


图 6-28 溴乙锭掺入 DNA。溴乙锭的插入增加了相邻碱基对之间的空间距离，扭曲了磷酸核糖骨架，降低了螺旋程度。

因为溴乙锭的结合增加了共价闭合环状（cccDNA）的 *Wr*，从而使 DNA 在凝胶电泳中的迁移率发生了很大变化。在非饱和的溴乙锭存在的情况下，负超螺旋环状 DNA 更为松弛并且迁移缓慢，而松弛的 cccDNA 则变为正超螺旋，迁移的速度加快。

RNA 的结构

RNA 含有核糖和尿嘧啶，通常是单链

现在我们来看 RNA 和 DNA 的 3 个不同之处（图 6-29）。第一，RNA 骨架含有核糖，而不是 2′-脱氧核糖，也就是说在核糖的 C2′的位置上带有一个羟基。第二，在 RNA 中尿嘧啶（uracil）取代了胸腺嘧啶，尿嘧啶有着和胸腺嘧啶相同的单环结构，但是缺乏 5′甲基基团。胸腺嘧啶实际上是 5′甲基-尿嘧啶。第三，RNA 通常是单链的多聚核苷酸。除了某些病毒以外，RNA 并不是遗传物质，它不需要以自己为模板进行复制。从功能上讲，mRNA 是从基因到蛋白质合成的媒介，RNA 其 tRNA 的另一功能是作为 mRNA 上密码子与氨基酸的“适配器”，同时，RNA 也是核糖体的组成部分；另外，RNA 还是一种调节分子，通过和某些 mRNA 互补序列的结合而对其翻译过程进行调控；最后，还有一些 RNA（包括组成核糖体的一种结构 RNA）是在细胞中催化一些重要反应的酶。综上所述，RNA 只是拷贝了双链 DNA 中的一条链，这条链的互补链则不存在。RNA 也可以形成很长的双螺旋，但这种情况在自然界中很少见。

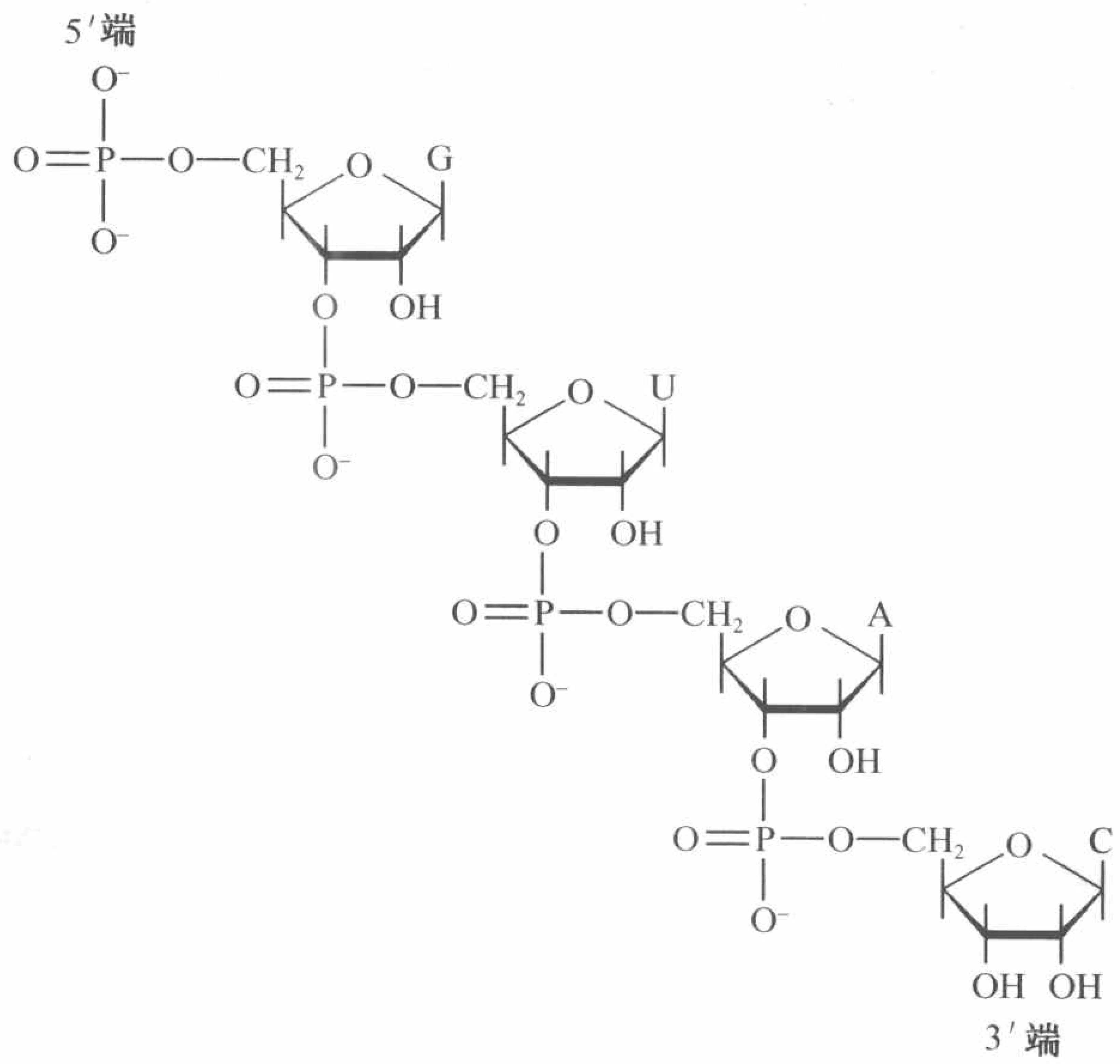


图 6-29　RNA 的结构特点。本图显示了 RNA 骨架的结构，由磷酸和核糖交替排列组成。RNA 结构中有别于 DNA 的部分以红色表示。

RNA 链自身折叠形成局部双螺旋，类似 A-DNA

尽管 RNA 是单链分子，它仍然具有大量的双螺旋结构的特征（图 6-30）。这是因为 RNA 链频繁发生自身折叠，从而在互补序列间形成碱基配对区。如果两段互补序列彼此非常接近，RNA 可以采用多种柄-环结构（stem-loop structure）中的一种，其中的非互补 RNA 以发夹、凸出或简单的环状形式从双螺旋区域的末端突出来。

某些情况下环的特殊性质能增强柄环结构的稳定性。例如，具有 UUCG“4 碱基环”的柄环结构由于环中的特殊的碱基堆积作用而非常稳定（图 6-31）。碱基配对也可以发生在不相邻的序列中，形成称为“假结”（pseudoknot）的复杂结构（图 6-32）。RNA 的碱基配对区可以是规则的双螺旋，或者是不连续的，例如，含有不互补的，凸出于双螺旋之外的核苷酸。

RNA 具有额外的非 Watson-Crick 配对的碱基，如 G∶U 碱基对，这个特征使 RNA 更易于形成双螺旋结构。G∶U 碱基对之间有两个氢键：尿嘧啶 N3 和鸟嘌呤 C6 上羰基之间的氢键，以及尿嘧啶 C2 上羧基和鸟嘌呤 N1 之间的氢键（图 6-33）。由于 RNA 链中不仅存在 4 种常规 Watson-Crick 碱基对，而且有 G∶U 碱基对，因

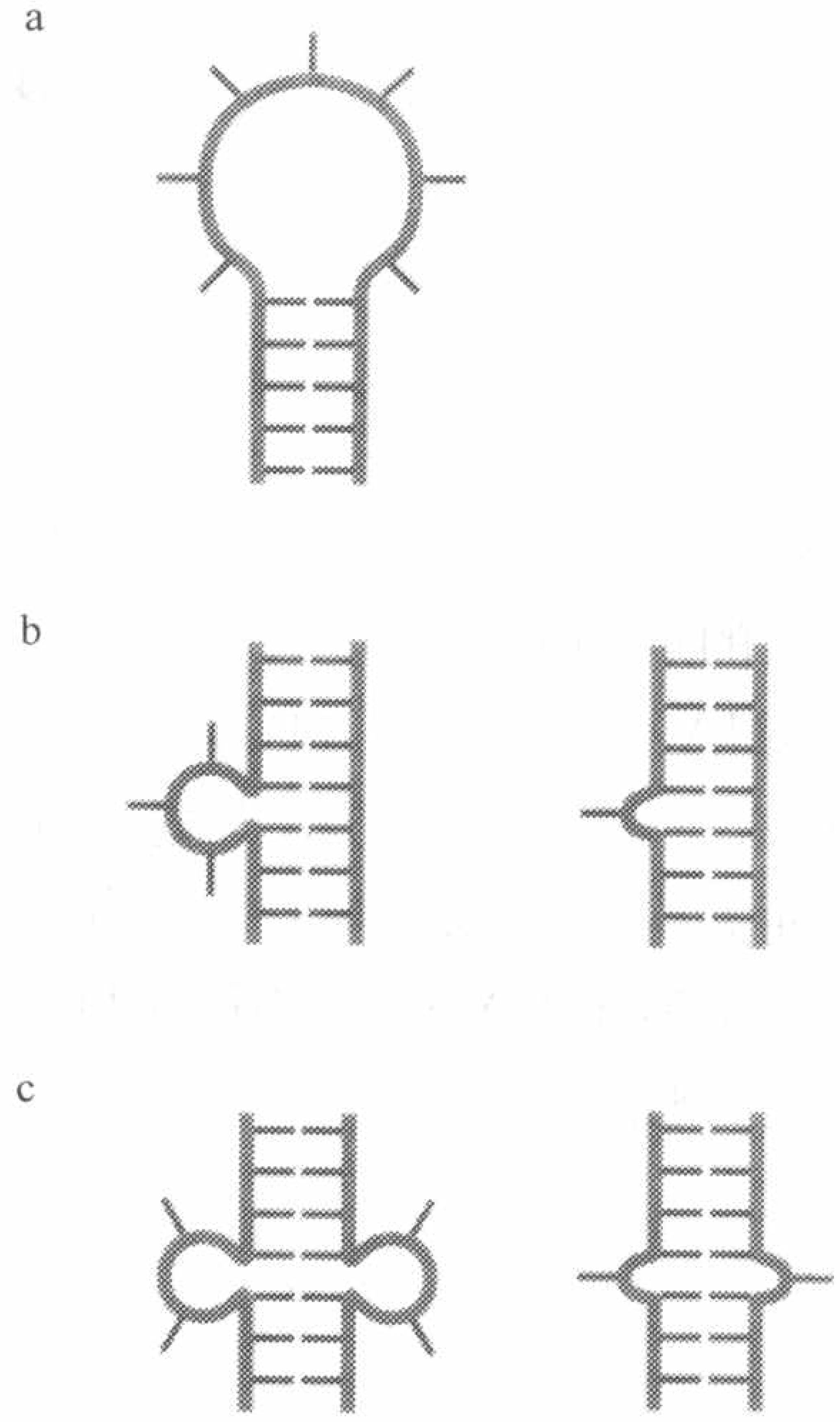


图 6-30 RNA 的双螺旋特点。在含有互补序列的 RNA 分子中，非互补的部分可以环状凸出出来，形成图示结构中的一种。a. 发夹结构。b. “凸”结构。c. 环。

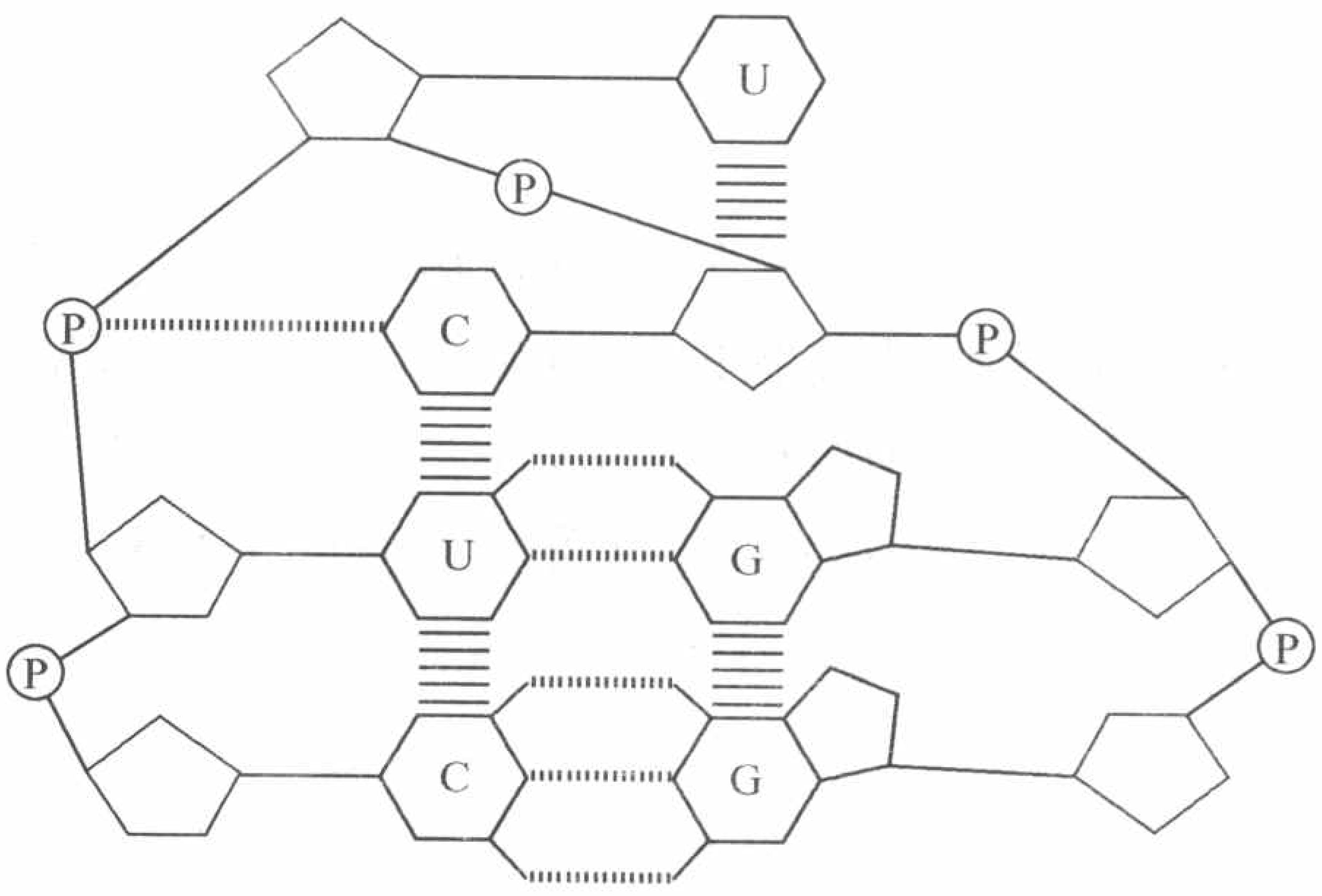


图 6-31 四碱基环。碱基堆积作用使四碱基环结构更加稳定。紫色的核糖之间的灰色的环代表 RNA 骨架的磷酸部分。水平线代表碱基堆积作用。

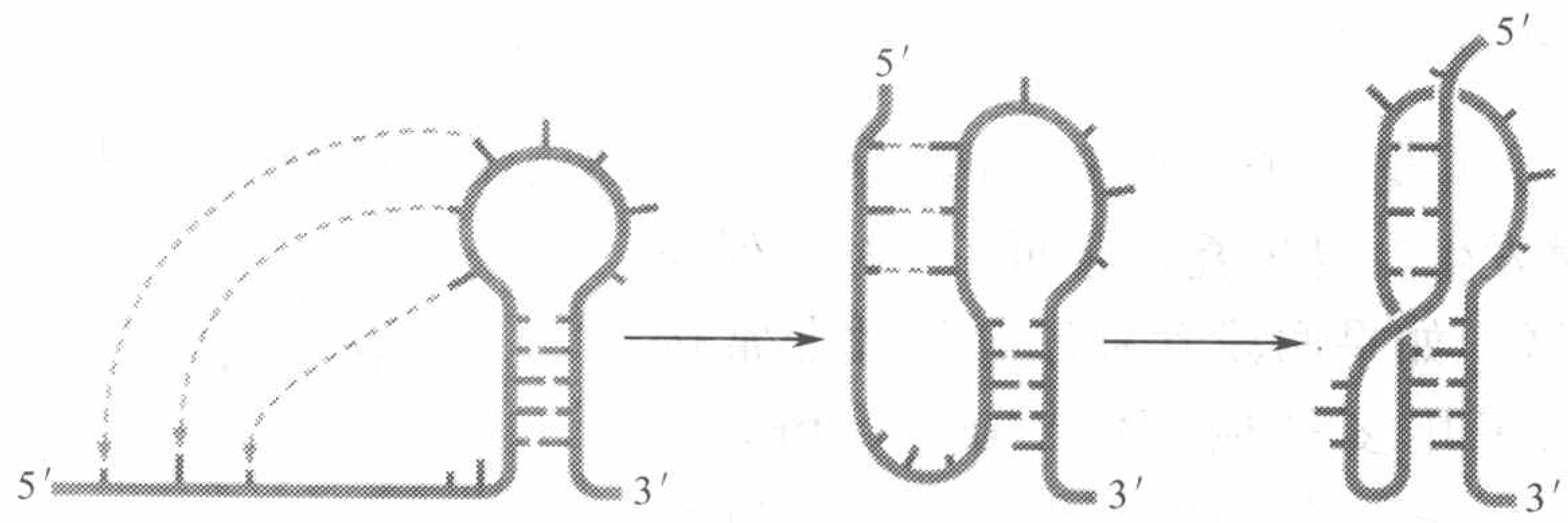


图 6-32　假结。不相邻的互补序列之间碱基配对形成假结结构。

此，RNA 链自我配对的能力得到增强，常常出现局部区域碱基配对，但不是 DNA 那种长长的规则的螺旋形式。

RNA 骨架上 2′-OH 的存在阻止 RNA 形成 B 型螺旋。其实双螺旋 RNA 更类似于 A-DNA 的结构。它的小沟宽且浅从而易于接近，但是小沟里几乎没有序列特异性信息；而大沟狭且深，与它相互作用的蛋白质的氨基酸侧链难以接近它。所以，从精细原子结构上看 RNA 的双螺旋与 DNA 的双螺旋十分不同，使 RNA 不适合与蛋白质进行序列特异性的相互作用（虽然有一些蛋白质可以序列特异性方式结合在 RNA 上）。

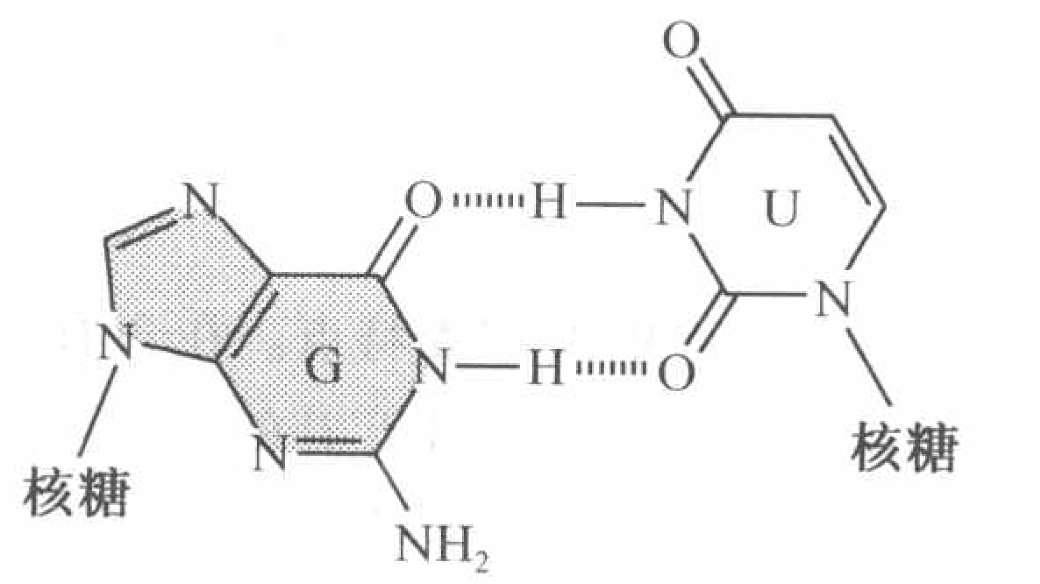


图 6-33　G∶U 碱基对。图中所示为鸟嘌呤与尿嘧啶配对形成两条氢键。

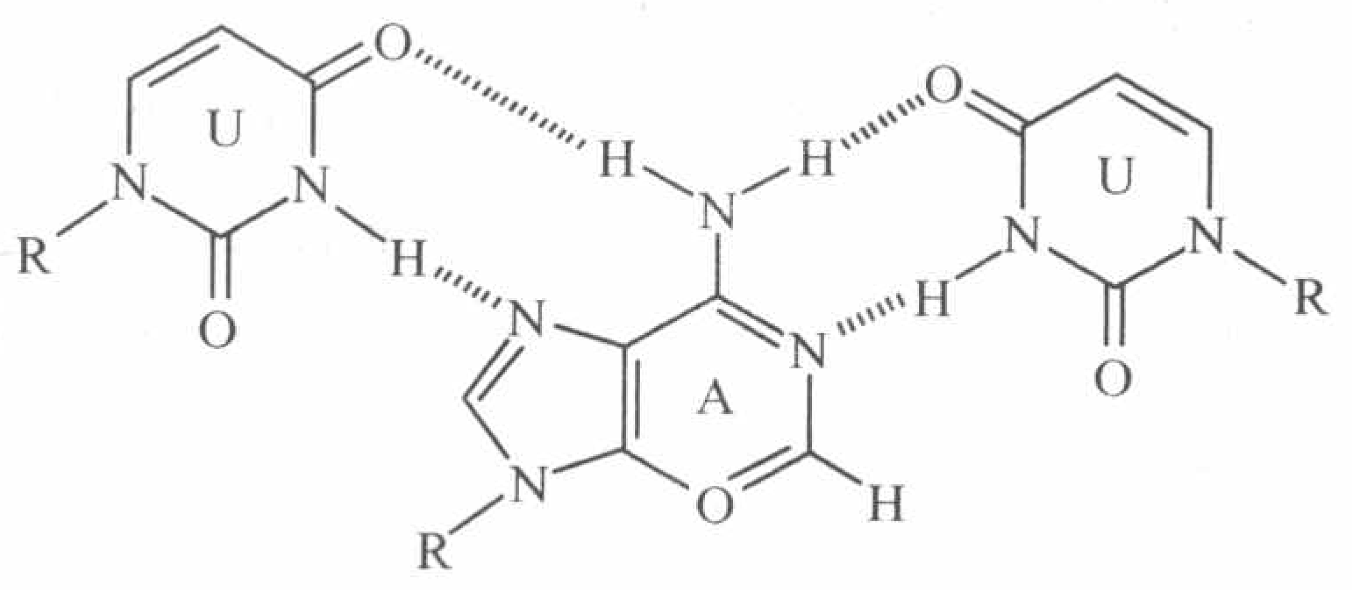


图 6-34　U∶A∶U 三碱基对。图中所示为非常规三碱基配对形成的氢键。

RNA 可折叠成复杂的三级结构

由于 RNA 没有形成长的规则螺旋的限制，因此可形成大量的三级结构。因为 RNA 骨架上未配对的区域可以不受限制地自由旋转，所以 RNA 常常折叠成包括不规则的碱基配对的复杂的三级结构。例如，tRNA 中的三碱基对以及碱基与骨架的相互作用（如图 6-34 中的 U：A：U 三碱基）。蛋白质可以协助 RNA 大分子三级结构的形成，如核糖体中发现的蛋白质。蛋白质可以中和磷酸骨架的负电荷，从而消除由于静电斥力造成的结构不稳定。

研究者利用 RNA 的潜在结构复杂性创造出新的 RNA 种类（自然界中不存在），使之具有符合人们需要的特性。以随机序列合成 RNA 分子可能产生出寡聚核苷酸的混合物，表现了序列的多样性。举例来说，序列的每个位置可能是 4 种核苷酸中的一种，因此长度为 20bp 的寡聚核苷酸的混合物就有 4^{20} 或是 10^{12} 种可能性。从混合物中以生化方法可以选出具有独特性质的 RNA 分子，如与特定小分子具有亲和力的 RNA。

一些 RNA 可以是酶类

许多年来普遍认为酶只能是蛋白质，酶必须能结合底物，启动化学反应，释放产物，并可如此重复多次。蛋白质很适合作为酶分子，它由 20 种氨基酸组成，可折叠形成具有底物、小分子辅助因子的结合位点和催化位点的复杂三级结构。现在我们知道很容易形成三级结构的 RNA 也是一种生物催化剂。RNA 酶被称为核酶（ribozyme），具有典型酶的许多特性：催化位点、底物结合位点和辅助因子（如金属离子）结合位点。

最早被发现的核酶是 RNase P，一种核糖核酸酶，参与从大的 RNA 前体生成 tRNA。RNase P 由 RNA 和蛋白质共同组成，起催化作用的只是 RNA。蛋白质组分中和了 RNA 上的负电荷，使 RNase P 可与带负电的底物有效结合，从而促进了反应的进行。如果有小的带正电的离子（如多肽亚精胺）中和了负电荷，即使没有蛋白质组分，RNA 组分也可以催化 tRNA 前体的剪切。mRNA 前体、tRNA 前体和 rRNA 前体间插序列即内含子（intron）的切除过程称为 RNA 剪接（RNA splicing）（第 13 章），其他核酶在这个过程中催化所谓的转酯反应。

锤头状核酶通过形成 2′，3′-环磷酸剪切 RNA

在结束 RNA 的讨论之前，让我们再关注一下一种特殊核酶的结构及功能，即所谓的锤头状核酶（hammerhead）（见结构指南 6-2）。它是一种序列特异性的核糖核酸酶，从称为类病毒（viroid）的植物感染性 RNA 病毒中得到，其依赖自我剪切得以繁殖。当类病毒复制时，在连续的 RNA 链上产生多个拷贝。单个类病毒由剪切而产生，而这种剪切反应则由结合处周围的 RNA 序列催化完成。这样的自我剪切的序列由于其二级结构类似锤头而被称为锤头结构，而每个锤头结构由 3 个配对的柄构成（Ⅰ、Ⅱ、Ⅲ），包围着起催化作用的不互补的核苷酸中心（图 6-35）。锤头结构的三级结构更接近 Y 字型（图6-36）。

a

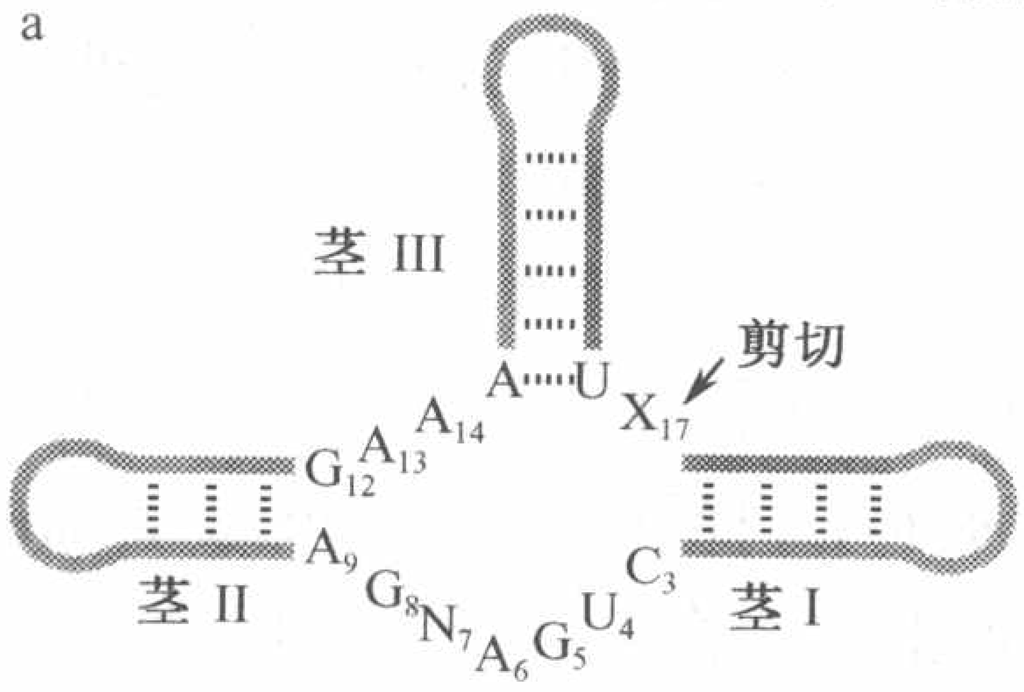


图 6-35 锤头状核酶的二级结构。分子结构显示了每个柄的两部分是如何通过环而连接在一起的，但是 3 个柄并不必都成环。事实上，在类病毒里，柄Ⅲ的两个部分并没有被环连接起来。a. 锤头状核酶的预测二级结构。Watson-Crick 碱基对的相互作用被标记成红色；红色箭头指向将断裂的键；可能为最小的底物链被标记为蓝色；（U）尿嘧啶，（A）腺嘌呤，（C）胞嘧啶，（G）鸟嘌呤。[来源：a. 经 McKay D. B. and Wedekind J. E. 1999，The RNA World，2nd edition（ed. R. F. Gesteland et al.）p. 267，Fig. 1A. ©Cold Spring Harbor Laboratory Press.]

为了弄清锤头结构如何起作用，必须首先了解在碱性环境下 RNA 如何水解。如果 pH 很高，RNA 骨架上核糖的 2′羟基会处于去质子化状态，这样使带有负电的氧可以进攻位于同一核糖的 3′易断开的磷，从而打断 RNA 链，产生2′,3′-环磷酸酯以及一个自由的 5′羟基。RNA 链上的每一个核糖都能通过这个反应水解，这样整个完整的分子能够酶解成核苷酸分子。（想一想为什么 DNA 不像 RNA 那样易于被碱水解?）许多蛋白质类的核糖核酸酶也是通过2′,3′-环磷酸酯的形成来裂解 RNA 底物。在正常细胞环境中的 pH 时，这些蛋白质酶

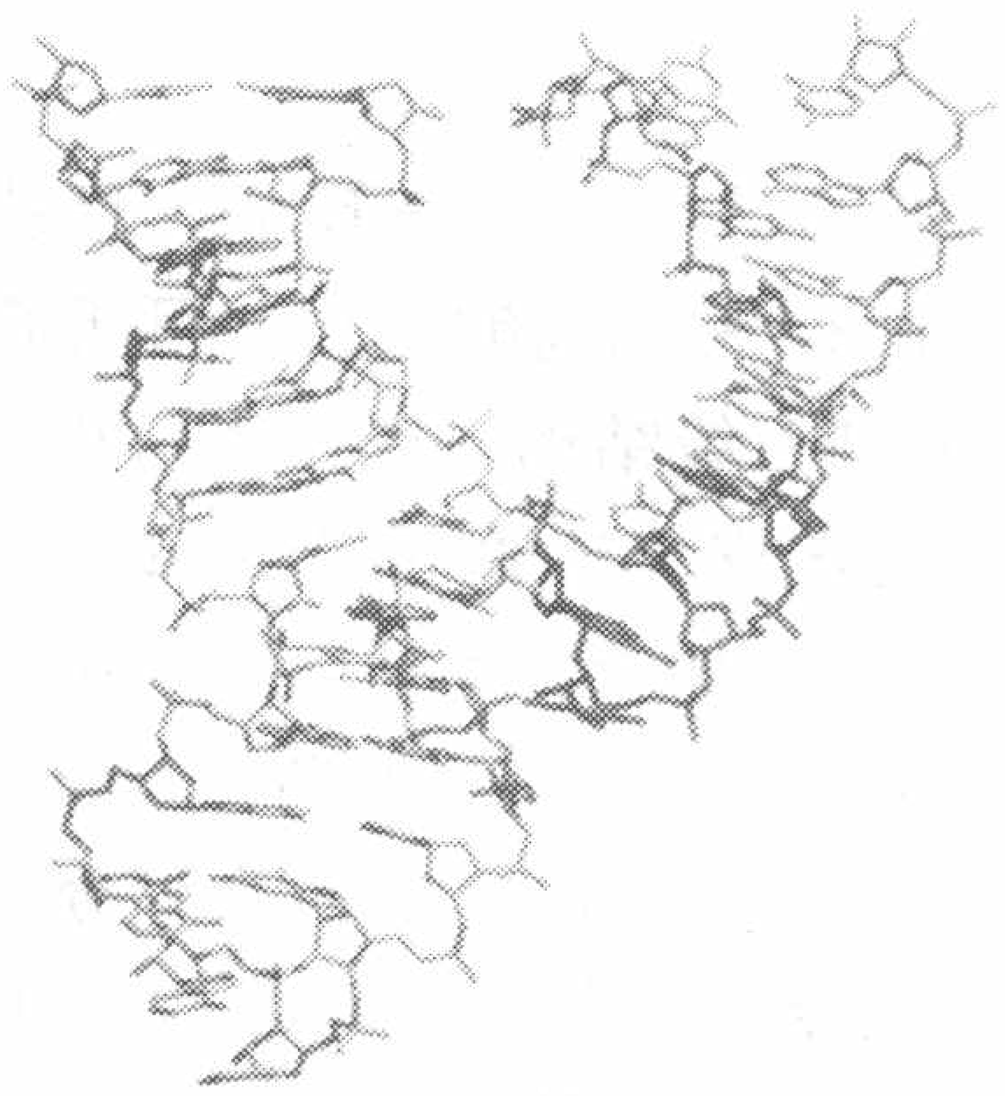


图 6-36 锤头状核酶的三维结构。图中显示的是锤头状核酶的精细结构：柄Ⅰ（右上），柄Ⅱ（左上）以及柄环Ⅲ（底部）。红色表示位于胞嘧啶（位置为 17）处的剪切位点（见图 6-35）。（ScottW. G.，Finch J. T.，and Klug A. 1995，Cell 81：991，图片用 MolScript，BobScript 及 Raster3D 软件制作）

类通过结合在其活性位点上的金属离子来活化 RNA 的 2′羟基。锤头状核酶同样是通过2′,3′-环磷酸酯的形成来裂解 RNA 底物，但裂解反应的机制还有待进一步研究。

因为锤头状核酶通常的作用方式是自身剪切，它们不能算是真正的催化剂。因为每个催化剂分子通常一次只催化一个反应，这样转换数为一。但是可以通过遗传工程手段把整个分子分开成两部分，一部分是核酶，包含了催化核心；另一部分是底物，包含了剪切位点。底物同核酶的柄Ⅰ和柄Ⅲ（图 6-35）结合。剪切完成之后底物被释放，代替它的是一个没有被剪切的新底物，因此可以进行重复的剪切。

生命是否起源于 RNA 世界？

核酶的发现深刻的改变了我们对于生命可能起源的看法。我们现在可以设想也许曾经存在全部以 RNA 为基础的原始生命形式。在这种生命世界里，RNA 既是遗传物质又是酶。我们知道现在的生命形式进行信息传递的时候需要有 DNA、RNA 和蛋白质，而这种单一的 RNA 生命形式更早。有发现表明蛋白质世界可能起源于 RNA 世界。核糖体中负责形成肽键的组分，即肽酰转移酶就是 RNA 分子（第 14 章）。与 RNase P 的锤头状核酶和其他已经发现的核酶都通过作用于磷中心而产生肽键不同，肽酰转移酶却是通过作用于碳中心而催化这个反应。这一发现把 RNA 的化学性质同蛋白质世界里最基础的化学反应——肽键形成联系起来。也许核糖体核酶正是从原始生命残留下来的，那个时候所有的酶都是 RNA。

小结

DNA 通常形成右手双螺旋。两条多聚脱氧核糖核苷酸链构成双螺旋。每一条链又是由交替的脱氧核糖和磷酸以磷酸二酯键依次连接形成的多聚体。每个脱氧核糖都会伸出一个碱基，碱基有 4 种：属于嘌呤类的腺嘌呤和鸟嘌呤，属于嘧啶类的胸腺嘧啶和胞嘧啶。脱氧核糖-磷酸骨架是规则的，而碱基的顺序则是不规则的，使其承载了 DNA 的信息。每一链都有 5′→3′的方向性，两条链按照反向平行的方式对应，即两条链的方向相反。

碱基之间的配对使两条链结合到一起。碱基通过氢键特异配对：一条链上的腺嘌呤总是和另一条链上的胸腺嘧啶配对，而鸟嘌呤则通常和胞嘧啶配对。这种碱基之间的严格配对表明构成 DNA 碱基的嘌呤和嘧啶的氢原子位置是固定的。腺嘌呤和胞嘧啶以氨

基而不是亚氨基异构体形式存在，而鸟嘌呤和胸腺嘧啶则以酮式而不是烯醇式形式存在。两条链间的碱基互补赋予了 DNA 自我编码的特性。

DNA 双螺旋两条链的分开（变性）需要高温、极端的 pH 或者其他某些变性剂来破坏氢键。当 DNA 缓慢地恢复到正常的细胞环境，变性的双链会特异恢复到有生物活性的双螺旋结构（复性或退火）。

DNA 在溶液状态下周期性螺旋缠绕，每旋转一周有大约 10.5 个碱基。碱基对相互之间的堆积形成了带有两条沟的双螺旋。由于糖环在双螺旋外侧与碱基之间形成 120°的夹角，使两条沟的大小不同。每个碱基对的边缘暴露在沟内，形成富含氢键供体和受体及范德华表面的区域，可用于识别碱基对。大沟（major groove）包含的化学信息比小沟（minor groove）更多，而且对核酸序列专一性结合蛋白的识别作用更为重要。

几乎所有细胞内的 DNA 都是非常长的分子，每条特定的染色体就是一个 DNA 分子。真核细胞容纳如此超长的 DNA 采用的方法是利用蛋白颗粒把 DNA 分段包装成核小体。大多数 DNA 分子以线形存在，但也有的是环状的，如原核细胞和某些病毒的染色体。

DNA 分子同样具有柔性。如果其拓扑结构不受限，它可以自由旋转从而改变两条链相互缠绕的次数。但是当它以闭合共价环状结构存在或者以染色质存在时，DNA 的拓扑结构就会受到限制。连环数是闭合环状 DNA 的固有拓扑学特性，它是指如果需要把两条环状的链完全解开，一条链需要绕过另一条链的次数。连环数是两个可以相互转换的几何特性参数的总和：扭转数，表示两条链相互缠绕的次数；缠绕数，表示 DNA 在长轴上自身交叉的数目。DNA 每个螺旋 10.5 个碱基并且螺旋没有扭曲，此时 DNA 在生理条件下处于松弛状态。如果连环数减少 DNA 受到扭转压力时候的状态称为负超螺旋。细胞中的 DNA 一般有 6%以负超螺旋存在。

真核细胞中以左手方向缠绕核小体的 DNA 引入了负超螺旋；而原核生物由于没有组蛋白所以负超螺旋的产生是由促旋酶来完成的。促旋酶是拓扑异构酶Ⅱ家族中的一种，拓扑异构酶Ⅱ在 DNA 双链上产生瞬间的断裂，然后使双链的一个区域穿过缺口，通过这两步可改变 DNA 的连环数。拓扑异构酶Ⅱ中的一些成员能够起到松弛超螺旋 DNA 的作用，相反，DNA 促旋酶能够产生负超螺旋。拓扑异构酶Ⅰ也能起到松弛超螺旋 DNA 的作用，但是它的作用方式是使一条链穿过另一条链瞬间产生的缺口。

RNA 与 DNA 相比有以下不同：它的骨架构成是核糖而不是 2′-脱氧核糖；尿嘧啶代替了胸腺嘧啶的位置；它通常以多核苷酸单链形式存在，没有互补链。由于是单链 RNA 在自身互补区域可以自我折叠形成局部双螺旋。与 DNA 相比，RNA 的碱基配对更宽。除了 A∶U 配对和 C∶G 配对，U 和 G 也可以配对。这一特性使 RNA 能够形成不符合 Watson-Crick 碱基配对原则的碱基对，从而增大了 RNA 形成局部双螺旋的能力。由于没有形成长的规则的双螺旋的限制，碱基和核糖磷酸骨架之间的非常规相互作用使 RNA 能形成复杂的三级结构。

有些 RNA 同时也是一种酶——它们能够催化细胞内和离体的化学反应，这些 RNA 酶被称为核酶。大多数核酶作用于磷中心，如核糖核酸酶 RNase P。RNase P 由蛋白质和 RNA 构成，但是，只有它的 RNA 部分才起催化作用。锤头状核酶是具有催

化自身断裂作用的 RNA，通过产生2′,3′-环磷酸酯来切开 RNA 骨架。肽基转移酶是另一类核酶，作用于碳中心。肽基转移酶催化肽键的形成，是组成核糖体的 RNA 组分之一。对作用于磷中心或碳中心的核酶的发现，表明在生命起源的早期，RNA 既是遗传物质，又起到了酶的作用。

（李京湘　译　朱金桂　胡松年　刘韧　杨焕明　校）

参考文献

书

Bloomfield V.A., Crothers D.M., Tinoco I., Jr., and Heast J.E. 2000. *Nucleic acids: Structures, properties, and functions.* University Science Books, Sausalito, California.

Cold Spring Harbor Symposium on Quantitative Biology. 1982. Volume 47: Structures of DNA. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Gesteland R.F., Cech T.R., and Atkin J.F., eds. 1999. *The RNA world,* 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

DNA的结构

Chambers D.A., ed. 1995. DNA: The double-helix—Perspective and prospective at forty years. *Ann. N.Y. Acad. Sci.* **758:** 1–472.

Dickerson R.E. 1983. The DNA helix and how it is read. *Sci. Am.* **249:** 94–111.

Franklin R.E. and Gosling R.G. 1953. Molecular configuration in sodium thymonucleate. *Nature* **171:** 740–741.

Roberts R.J. 1995. On base flipping. *Cell* **82:** 9–12.

Watson J.D. and Crick F.H.C. 1953. Molecular structure of nucleic acids: A structure for deoxyribonucleic acids. *Nature* **171:** 737–738.

———. 1953. Genetical implications of the structure of deoxyribonucleic acids. *Nature* **171:** 964–967.

Wilkins M.H.F., Stokes A.R., and Wilson H.R. 1953. Molecular structure of deoxypentose nucleic acids. *Nature* **171:** 738–740.

DNA拓扑学

Dröge P. and Cozzarelli N.R. 1992. Topological structure of DNA knots and catenanes. *Methods Enzymol.* **212:** 120–130.

Wang J.C. 2002. Cellular roles of DNA topoisomerases: A molecular perspective. *Nat. Rev. Mol. Cell Biol.* **3:** 430–440.

RNA的结构

Darnell J.E., Jr. 1985. RNA. *Sci. Am.* **253:** 68–78.

Doudna J.A. and Cech T.R. 2002. The chemical repertoire of natural ribozymes. *Nature* **418:** 222–228.

Doudna J.A. and Lorsch. J.R. 2005. Ribozyme catalysis: Not different, just worse. *Nat. Struct. Mol. Biol.* **15:** 394–402.

Nelson J.A. and Uhlenbeck O.C. 2006. When to believe what you see. *Mol. Cell* **23:** 447–450.

第7章 基因组结构、染色体、染色质和核小体

本章概要

- 基因组序列和染色体多样性
- 染色体的复制和分离
- 核小体
- 染色质的高级结构
- 染色质结构的调控
- 核小体的组装

我们在第 6 章中单独介绍了 DNA 的结构。然而，在细胞中 DNA 是和蛋白质结合存在的，每条 DNA 及其结合的蛋白质构成了一条染色体（chromosome）。在原核生物、真核生物，甚至病毒中，DNA 都是以这种结构存在。DNA 包装为染色体具有几个重要的功能：首先，染色体是 DNA 的紧密结构，更适合存在于细胞中；其次，这种包装可以保护 DNA 免于损伤。完全裸露的 DNA 分子在细胞中是相当不稳定的。相反，染色体 DNA 是非常稳定的；其三，只有包装成染色体的 DNA 才能在每次细胞分裂时有效地将 DNA 传递给两个子代细胞；最后，染色体将每个 DNA 分子全面地组织起来。这种组织有助于基因的表达和亲本染色体之间的重组，使所有生物的不同个体之间产生多样性。

真核细胞染色体分子质量的一半为蛋白质。在真核细胞中，DNA 的一段给定区域及与之结合的蛋白质称为染色质（chromatin），这些结合蛋白大部分是小的碱性蛋白质，即组蛋白（histone）；其他丰度较低的其核染色体结合蛋白，通常称为非组蛋白（non-histone protein）。这些蛋白质包含大量的 DNA 结合蛋白，可以调控细胞 DNA 的转录、复制、修复和重组。这些内容将在随后的 5 章中逐一详细讨论。

染色质中的蛋白质有另一重要的功能：压缩 DNA。下面的计算可以清晰地看出该功能的重要性。人的细胞中单套染色体（单倍体）含有 3×10^9 碱基对，我们在第 6 章里学到，每个碱基对的厚度为 3.4Å，如果将单倍体 DNA 拉直并首尾相接，DNA 的长度将约为 10^{10} Å，即 1m。对于一个二倍体细胞（人的细胞是典型的二倍体细胞）来说，DNA 的长度加倍为 2m。由于一个典型的人的细胞核的直径仅为 10～15μm，显而易见 DNA 必须被压缩多个数量级才能适应那么小的空间。这是怎样做到的呢？

在人的细胞（和其他所有的真核细胞）中，DNA 的压缩主要是通过 DNA 有规律地与组蛋白的结合，DNA 与组蛋白结合所形成的结构称为核小体（nucleosome）。压缩

结构的 DNA 长度线性只有原先的万分之一，而核小体结构的形成是允许 DNA 压缩为如此紧密结构的过程的第一步。当然，DNA 的压缩也会造成损失。DNA 与组蛋白及其他包装蛋白的结合限制了对 DNA 的可接近性，可接近性的降低会干扰介导 DNA 复制、修复、重组的蛋白质与 DNA 的相互作用，尤其是对转录相关的蛋白质影响最显著。真核细胞利用染色质的拟制特性和调控基因表达和其他 DNA 有关的生命活动。

个别核小体的结构变化可以使染色体特定区域的 DNA 与其他蛋白质相互作用。核小体的改变是由修饰和重塑核小体的酶所介导，这是个动态的、区域性的过程，可以使酶和调控蛋白在不同的时间结合到染色体的不同区域。因此，了解核小体的结构及其与 DNA 相关的调控是解释真核细胞中涉及 DNA 的多数过程的调控的关键。

尽管原核细胞的基因组都较小，但是对其 DNA 进行压缩仍是必要的。*E. coli* 必须把它的大约 1mm 的 DNA 包装到长度仅有 1μm 的细胞中。目前尚不清楚原核 DNA 是如何进行压缩的。例如，细菌没有组蛋白和核小体，却有其他小的碱性蛋白质，这些蛋白质可能会有相似的功能。本章我们将集中介绍真核细胞的染色体和染色质。首先，我们将讨论不同生物体染色体的 DNA 序列特征，尤其集中在蛋白质编码内容的变化上；随后，将讨论保证细胞分裂时染色体能精确地传递的机制；最后，我们将介绍真核染色质及其基本功能单元——核小体的结构和调控。

基因组序列和染色体多样性

在我们详细介绍染色体结构之前，有必要先了解构成染色体骨架的 DNA 分子的结构。近来对几千个生物基因组的测序提供了丰富的信息，使我们了解染色体 DNA 的构成及其特性是如何随着生物体复杂性的增加而变化的。

染色体可以是环状或线性的

传统的观点认为原核细胞有一条环状的染色体，真核细胞有多条线性染色体（表 7-1）。随着对越来越多的原核生物的研究，这个观点受到了挑战。尽管大多数研究过的原核生物（如 *E. coli* 和 *B. subtilis*）确实只有单一的环状染色体，然而现已发现有大量的原核生物含有多条染色体，或线性染色体，或二者皆有。相反，所有的真核细胞都含有多条线性染色体。真核生物染色体的数目因种而不同，通常在 2～50 条，但在极少的情况下染色体数目能达到数千条［例如原生生物四膜虫（*Tetrahymena*）的大核，表 7-1］。

表 7-1 不同生物体染色体组成的变化

物种	染色体数目	染色体拷贝数	染色体形式	基因组大小/Mb
原核生物				
生殖道支原体	1	1	环状	0.58
大肠杆菌 K-12	1	1	环状	4.6
根瘤农杆菌	4	1	3 条环状，1 条线性	5.67

续表

物种	染色体数目	染色体拷贝数	染色体形式	基因组大小/Mb
苜蓿中华根瘤菌	3	1	环状	6.7
真核生物				
酿酒酵母（芽殖酵母）	16	1 或 2	线性	12.1
粟酒裂殖酵母（裂殖酵母）	3	1 或 2	线性	12.5
秀丽线虫（线虫）	6	2	线性	97
拟南芥（草本植物）	5	2	线性	125
黑腹果蝇（果蝇）	4	2	线性	180
四膜虫（原生生物）	小核 5，大核 225	小核 2，大核 10～10 000	线性	125
红鳍东方鲀（鱼类）	22	2	线性	393
小鼠（啮齿类）	19+X，Y	2	线性	2600
人	22+X，Y	2	线性	3200

环状和线性染色体都有其各自的问题，必须克服它才能保证基因组正常的维持和复制。环状染色体复制后需要拓扑异构酶来将其分离到子代细胞中，没有这些酶，复制后两个子代染色体仍然会互锁或连接在一起（图 6-23），相反，线性真核染色体的 DNA 末端必须保护起来以避免酶对 DNA 末端的降解，这造成了 DNA 复制过程中的一系列不同的困难，我们将在第 8 章中进行讨论。

每个细胞都有特定的染色体数目

典型的原核细胞的染色体仅有一个完整拷贝，包装在一个称为类核（nucleoid）的结构中（图 7-1b）。然而，当原核细胞迅速分裂时，正在复制过程中的那部分染色体会存在 2 个有时甚至是 4 个拷贝。原核细胞中还经常携带一个或多个小的独立的环状 DNA，称为质粒（plasmid）。与大的染色体 DNA 不同，质粒通常不是细菌生长所必需的，而是携带着赋予细菌良好特性的基因，如抗生素抗性基因。此外，质粒有别于染色体 DNA 的另一特征是每个细胞中可以有多个完整拷贝存在。

大多数真核细胞是二倍体（diploid），也就是说，细胞中每条染色体有两个拷贝（图 7-1c）。同一个染色体的两个拷贝叫做同源染色体（homolog），分别来自父本和母本。但是，一个真核生物中的所有细胞并非都是二倍体，某些细胞是单倍体或多倍体。单倍体（haploid）细胞每条染色体只有一个拷贝，并且参与有性生殖（例如，精子和卵子都是单倍体细胞）。多倍体（polyploid）细胞中每条染色体都超过两个拷贝。事实上，许多生物体维持它们大多数成熟细胞在多倍体状态。在极端情况下，每条染色体的拷贝数可能有数百甚至数千。这种全基因组扩增的方式可以使一个细胞产生大量的 RNA 和随后的蛋白质产物。例如，巨核细胞（megakaryocyte）是一种特化的多倍体细胞（每条染色体大约有 28 个拷贝），能产生数千个人体血液所必需的没有染色体的血小板（每毫升血液中大约有 200 000 个血小板）。通过多倍体形式，巨核细胞维持高水平的新陈代谢，以生产大量的血小板。由于这么多条染色体的分离是非常困难的事，因此

多倍体细胞几乎总是处于停止分裂的状态。无论染色体数目的多少，真核染色体总是包含在由膜包围的称为细胞核（nucleu）的细胞器中（图 7-1c）。

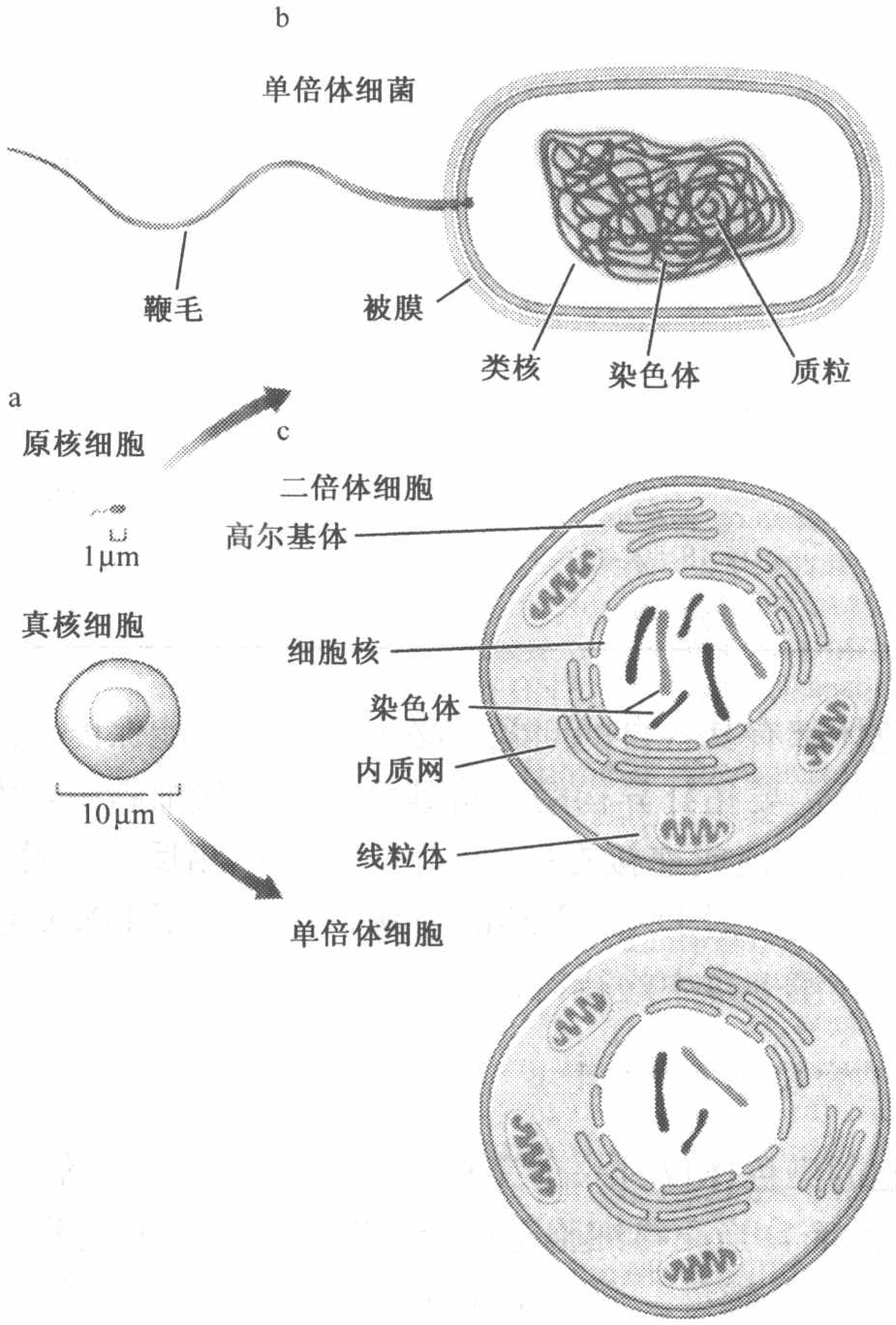


图 7-1 典型的原核细胞与真核细胞的比较。 a. 一个典型的真核细胞的直径约为 10μm，原核细胞约为 1μm。b. 原核染色体 DNA 位于类核内，占据细胞内大部分的空间。与真核细胞的细胞核不同，类核没有膜将它和细胞的其他组分分开，质粒 DNA 用红色表示。c. 真核染色体位于由膜所包围的细胞核中。由细胞核中存在的染色体的拷贝数可区分单倍体和二倍体细胞。（来源：Brown T. A. 2002 Genomes，2nd edition，p. 32. Fig2. 1. © BIOS Scientific Publishers. 经 Tayor & Francis 许可.）

基因组的大小与生物体的复杂度相关

基因组的大小（单倍染色体中 DNA 的长度）在不同的生物中有相当大的变化（表 7-2）。由于生物的复杂度越高所需的基因也就越多（比较细菌、单细胞真核生物和多细胞真核生物时可以看到这个规律，见 19 章）。因此基因组的大小与生物体的复杂度相关也就不足为奇。通常，原核细胞的基因组不足 10Mb。尽管一些较复杂的原生生物可能有大于 200Mb 的基因组，通常单细胞真核生物的基因组都小于 50Mb，而多细胞生物的基因组更大，甚至可超过 100 000Mb。

尽管生物体的复杂度与基因组的大小之间大致呈正相关，但并不精确。许多复杂度

相近的生物体其基因组大小却显著不同：果蝇的基因组大约是蝗虫的基因组的 1/25，而小麦的基因组是水稻基因组的 40 倍（表 7-2）。这些例子中，基因的数目而不是基因组的大小，与生物的复杂性更密切相关。当我们比较不同生物的相对基因密度时可以清楚地看到这一点。

表 7-2　不同生物体基因组中基因密度的比较

物　种	基因组大小/Mb	大致的基因数目	基因密度/(个/Mb)
原核生物（细菌）			
生殖器支原体	0.58	500	860
肺炎链球菌	2.2	2300	1060
大肠杆菌 K-12	4.6	4400	950
根瘤农杆菌	5.7	5400	960
苜蓿中华根瘤菌	6.7	6200	930
真核生物			
真菌			
酿酒酵母	12	5800	480
粟酒裂殖酵母	12	4900	410
原生生物			
四膜虫	125	>27 000	220
无脊椎动物			
秀丽线虫	103	20 000	190
果蝇	180	14 700	82
紫海胆（*Ciona intestinalis*）	160	16 000	100
东亚飞蝗	5000	未定	未定
脊椎动物			
河豚	393	22 000	56
人类	3200	20 000	6.25
小鼠	2600	22 000	8.5
植物			
拟南芥	120	26 500	200
水稻	430	约 45 000	约 100
玉米	2200	>45 000	>20
小麦	16 000	未定	未定
郁金香	约 120 000	未定	未定

E. coli 的基因组几乎全部由基因构成

E. coli 的单一染色体上的绝大多数 DNA 都编码蛋白质或结构 RNA（图 7-2）。而大部分非编码序列都参与调控基因的转录（第 16 章）。由于一个转录起始位点通常可以控制几个基因的表达，因而使这些区域在基因组中维持在最小的比例。*E. coli* 基因组中的一个关键元件并不是基因的一部分，也不是一段调控基因表达的序列，而是 *E. coli* 复制起始区。这个短的染色体区段指导着整个复制机制的组装（将在第 8 章讨论）。复制起始区虽然有重要的功能，但它仍是非常小的，仅为 *E. coli* 4.6Mb 基因组中的几百个碱基对。

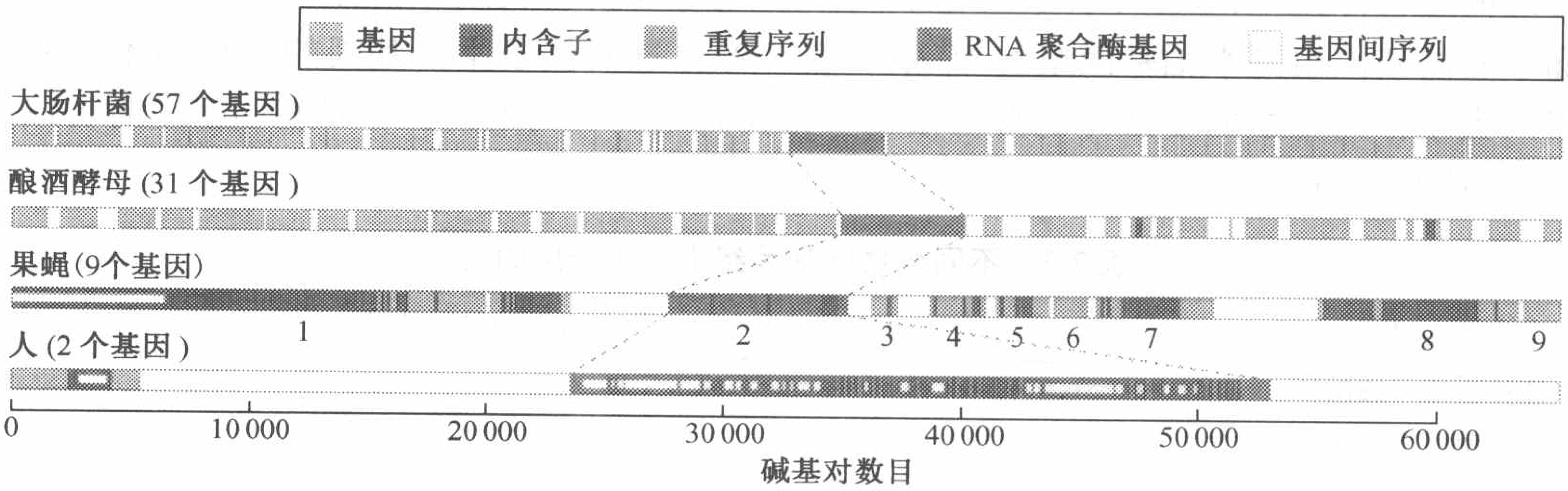


图 7-2 不同生物染色体基因密度的比较。图中展示了 4 种生物中一个具有代表性的 65kb 的 DNA 区段。编码 RNA 聚合酶（真核生物编码 RNA 聚合酶Ⅱ）大亚基的区域用红色表示，编码其他基因的区域用绿色表示，内含子 DNA 用紫色表示，重复 DNA 用黄色表示，独特的绿基因间 DNA 用灰色表示。注意随着生物复杂度的增加，在相同长度的 DNA（65kbDNA 区域）中编码基因的数目是如何变化的。

复杂度高的生物基因密度低

如何解释具有相似复杂度的生物体却有明显不同的基因组大小（如果蝇和蝗虫)?这种不同主要与基因的密度有关。基因密度的简单衡量方法是每 Mb 基因组 DNA 上基因的平均数目。如果一个 50Mb 基因组的生物有 5000 个基因，那么这个生物体的基因密度为 100 个/Mb。当比较不同生物体的基因密度时，可以清楚看到不同生物体利用 DNA 来编码可能的基因的效率不同。生物体的复杂度与基因的密度之间存在大致负相关的关系；生物体复杂度越低，其基因密度越高。例如，目前所发现的基因密度最高的生物体是病毒。在某些情况下病毒利用 DNA 的两条链来编码相互重叠的基因。尽管在细菌中重叠基因极少见，但细菌的基因密度非常恒定地接近 1000 个基因/Mb。

与原核生物相比，真核生物的基因密度较低且更多变（表 7-2)。在真核生物中仍然存在一个普遍的趋势，即基因密度随生物体复杂度的增加而降低。简单的单细胞真核生物酿酒酵母（*S. cerevisiae*）的基因密度（约 500 个基因/Mb）与原核生物非常接近。相反，人的基因密度估计是原核生物的 1/50。图 7-2 比较了在所有生物中相对保守的 RNA 聚合酶的大亚基基因的 DNA 序列，结果显示出不同的生物间基因大小存在极大的差异。较人基因组更大的生物可能有着更小的基因密度。是什么原因导致基因密度的降低呢?

基因仅占真核染色体 DNA 的一小部分

两种因素会导致真核细胞中基因密度的降低：基因大小和基因间序列（intergenic sequence）的增加。复杂度高的生物基因大的主要原因并不是蛋白质要大，或需要更长的 DNA 来编码相同的蛋白质，而是因为在真核生物中编码蛋白质的基因区域通常是不连续的。这些分散的非蛋白质编码区段称为内含子（intron）。内含子转录后可通过 RNA 剪接（RNA splicing）的过程从 RNA 中删除（图 7-3)，我们将在第 13 章详细讨论 RNA 剪接。内含子的存在极大地增加了编码一个基因所需的 DNA 的长度（表 7-3)。例如，人的一个基因的平均转录长度大约为 27kb（不要与基因密度相混淆)，而平均编码蛋白质的区段仅为 1.3kb。一个简单的计算可看出，人的编码蛋白质的基因中平均仅

有5%的序列直接参与了编码相应的蛋白质，而其他的95%都由内含子构成。简单的真核生物的内含子较少，这与它们较高的基因密度相一致。例如，在酿酒酵母中，仅有3.5%的基因有内含子，而且没有一个内含子大于1kb（表7-3）。

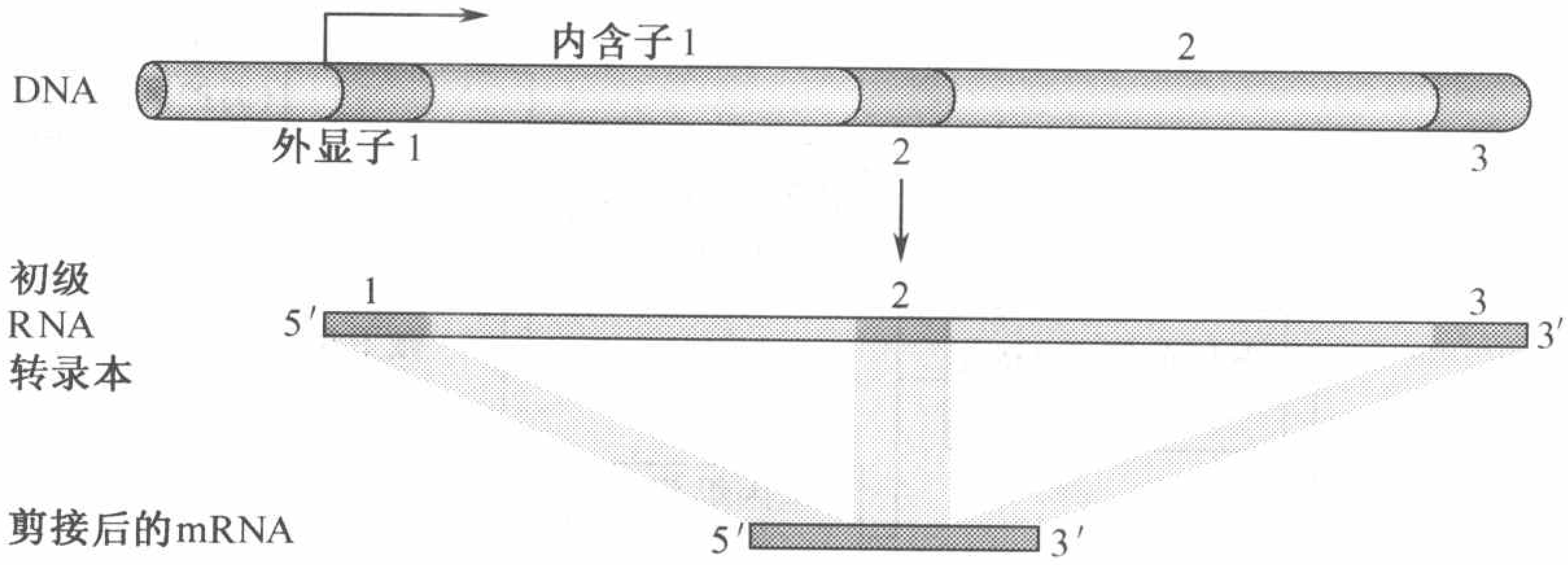


图7-3 RNA剪接的示意图。外显子1上方的箭头表示前体mRNA转录启动的位置。此初级转录本经过剪接加工去除非编码的内含子，生成信使RNA。

表7-3 不同基因组中内含子和重复序列的比例

物种	基因密度/（个/Mb）	基因中平均的内含子数目	重复序列的比例
原核生物（细菌）			
大肠杆菌K-12	950	0	<1
真核生物			
真菌			
酿酒酵母	480	0.04	3.4
无脊椎动物			
秀丽线虫	190	5	6.3
果蝇	82	3	12
脊椎动物			
河豚	56	5	2.7
人	6.25	6	46
植物			
拟南芥	220	3	未定
水稻	约100	未定	42

在更复杂的生物中基因间序列数目的大量增加是导致基因密度降低的另一原因。基因间隔区是基因组上与蛋白质和结构RNA的表达无关的那部分DNA。人的基因组中有超过60%的部分由基因间序列组成，这其中绝大多数的功能未知（图7-4）。基因间隔区DNA分为两类：单一的和重复的。大约1/4的基因间隔区DNA是单一的，导致单一基因间DNA序列增加的原因之一是指导和调控转录的DNA序列的增加，也称为调控序列（regulatory sequence）。随着生物体复杂度的提高，编码更多的基因，调控基因表达的调控序列也变得更加复杂，其序列大小也不断增加。人类基因组基因间DNA的单一区段包含了许多显而易见的已丧失功能的残留区段，如无功能的突变基因、基因碎段和假基因。突变基因和基因碎段是由于简单的随机突变或在DNA重组过程中的错误所造成。假基因则是因反转录酶（reverse transcriptase）的作用而形成（图7-5和第11章）。这种酶可将RNA复制为双链DNA（称为拷贝DNA，copy DNA或cDNA），通常只在某些特定类型的病毒中表达，参与病毒的繁殖。但是，当细胞被这类病毒感染

后，作为其副效应，细胞中的 mRNA 会被拷贝成 DNA，而这些 DNA 片段可以一个较低的频率重新整合到基因组中。然而，由于这些 DNA 缺少指导其表达的正确序列（这些序列一般不存在于基因的 RNA 产物中，见第 12 章），因此它们是不能表达的。

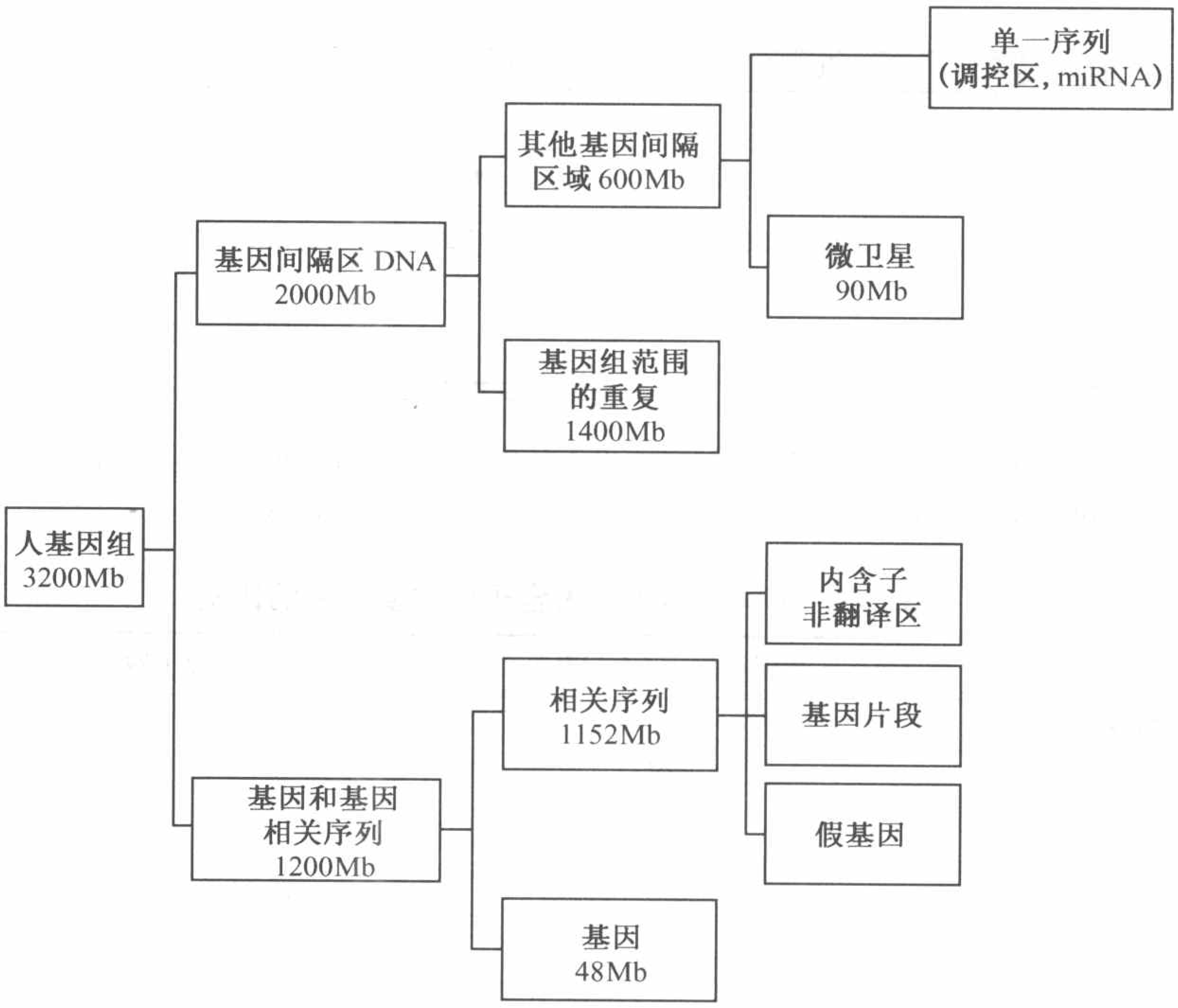


图 7-4 人类基因组的组织及其组分。人类基因组由许多不同类型的 DNA 序列构成，绝大多数 DNA 序列不编码蛋白质。此图显示了各种类型序列的分布和数量。（来源：Brown T. A. 2002. Genomes，2nd edition，p. 23，box 1. 4. © BIOS Scientific Publishers. 经 Taylor & Francis 许可.）

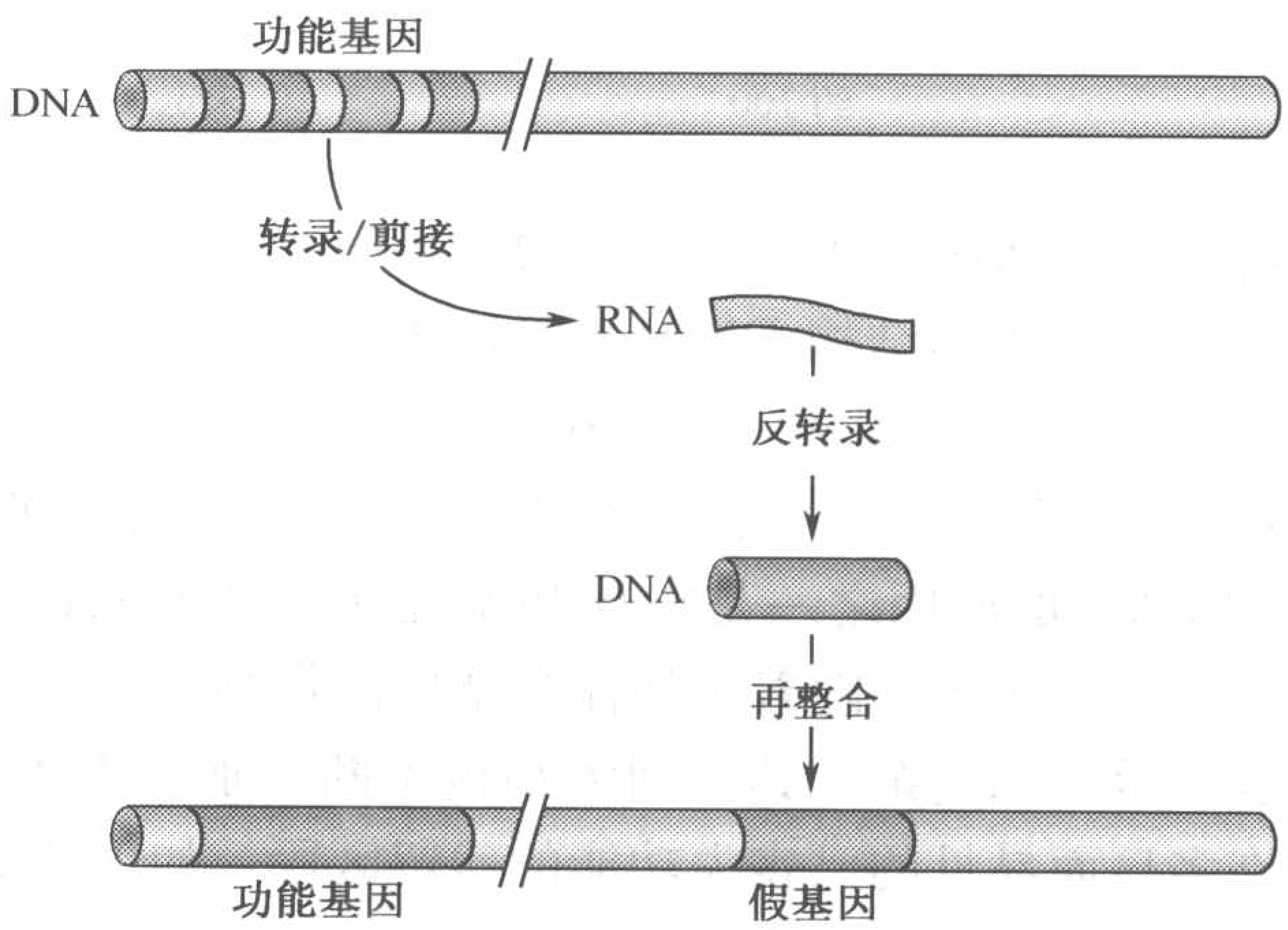


图 7-5 反转录 mRNA 的整合导致假基因的产生。当细胞中存在反转录酶时，mRNA 分子能被拷贝成双链 DNA。在极少数情况下，这些 DNA 分子能整合入基因组形成假基因。由于在新转录的 RNA 中内含子会被迅速地删除，所以假基因共同的特征就是缺少内含子。这可以将其与原基因区分开来。另外，因为启动子序列不是 mRNA 的一部分，所以假基因缺少相应的指导其转录的启动子序列。

最后，尽管我们对真核细胞单一基因间区域的功能还不清楚，但可以确定的是它在真核细胞中是有功能的，如最新发现的微 RNA（miRNA）。这些结构性小 RNA 通过改变 mRNA 的稳定性或者 mRNA 的翻译能力来调节其他基因的表达（参见微 RNA 的基因调控，第 18 章）。由于这些序列在最近才被发现，所以还没有包括在表 7-2 中。然而，在人类细胞中据估计有 400 多个 miRNA。多数单一性基因间区的另一功能还可能与复制的起始相关。尽管在多数真核生物中还没有被证明。

人的基因间隔区序列主要由重复 DNA 构成

几乎一半的人类基因组由在基因组中重复多次的 DNA 序列组成。重复 DNA 一般有两种：微卫星 DNA 和基因组范围的重复序列（genemo-wide repeat）。微卫星 DNA（microsatellite DNA）由非常短的（小于 13bp）串联重复序列构成。最常见的微卫星序列是二核苷酸重复序列（例如，CACACACACACACACA）。这种重复来自于 DNA 的正确复制的困难，几乎占了人类基因组的 3%序列。

基因组范围的重复序列要比微卫星大得多，每个重复单元的长度都大于 100bp，许多更是大于 1kb。这些序列要么作为单一拷贝散布在基因组中，要么紧密相连成簇存在。尽管这些重复有许多种类，但是它们有着共同的特点，即都是转座因子（transposable element）。

转座因子是指能从基因组的一个位置转移到另一个位置的序列，这个过程称为转座（transposition）。在转座过程中，转座元件经常在保留原拷贝的情况下，转移到基因组中新的位置上。因此，这些序列可在基因组中加倍并累积。在人的细胞中转座因子的移动极少发生。然而，在漫长的生命进化过程中，这些转座因子非常成功地繁衍了自己，以至于现在大约占人基因组的 45%。在第 11 章中我们将讨论转座因子在基因组中移动的机制以及如何控制它们的移动以防止它们整合到基因之中。

虽然我们是以人类基因组为背景讨论了基因间序列的特征，然而在其他生物体中也发现了许多相同的特征。例如，通过对基因组非常大的几种植物（如玉米）的部分已知序列的比较，可以看到在这些基因组中转座因子可能占有更大的比例。甚至在像 *E. coli* 和酿酒酵母这样基因组结构致密的生物中，也发现有转座因子和微卫星重复的存在（图 7-2）。不同之处只是这些序列在这些简单生物体中占据基因组不那么成功，这一不成功可能是由低效率的复制和（或）高效率的删除（被修复机制删除或是发生复制的生物体被删除）所综合影响的结果。

尽管现在倾向于认为重复 DNA 是垃圾 DNA（junk DNA），但是这些序列可以稳定地维持成百上千代，表明基因间 DNA 对宿主生物体有积极的意义（或者选择优势）。

染色体的复制和分离

真核染色体在细胞分裂过程中需要着丝粒、端粒和复制起始位点

在真核染色体上有几个重要的 DNA 元件，它们不是基因也不参与基因表达的调控

(图 7-6)。这些元件包括指导染色体 DNA 复制起始的复制起始位点(origins of replication)、引导染色体进入子代细胞的着丝粒(centromere)及负责保护和复制线性染色体末端的端粒(telomere)。它们对于细胞分裂过程中染色体的正确复制和分离是至关重要的。现在我们对每个元件逐一加以讨论。

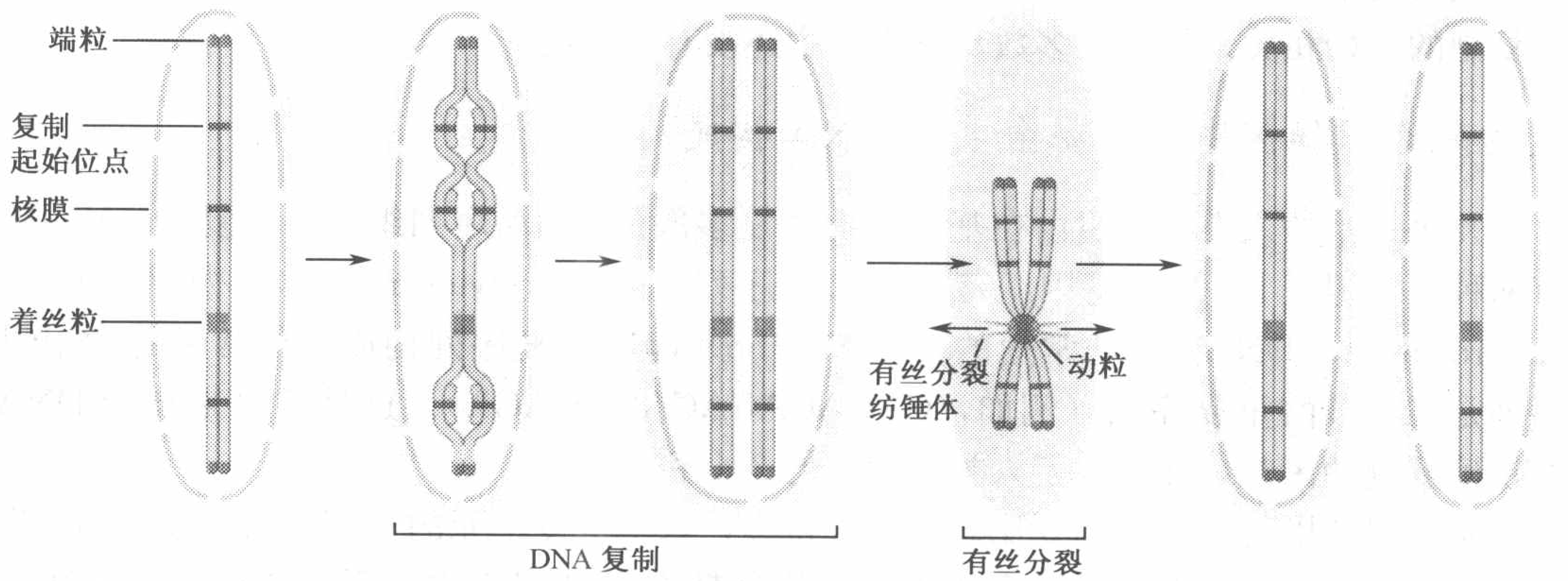


图 7-6 着丝粒、复制起始位点和端粒是真核染色体的稳定所必需的。每条真核染色体包括两个端粒、一个着丝粒和多个复制起始位点。端粒位于每条染色体的两端。与端粒不同,每条染色体上只有一个着丝粒,且没有固定的位置。有些着丝粒靠近染色体中部,有些靠近端粒。复制起始位点分布于整条染色体上(在酿酒酵母中大约每 30kb 有一个复制起始位点)。

复制起始位点是 DNA 复制机器组装和复制起始的位点。复制起始位点每隔约 30~40kb 分布在每条真核染色体上。原核染色体也需要复制起始位点。与真核染色体不同,原核染色体通常只有一个复制起始位点。一般来说,复制起始位点位于非编码区。复制起始位点的 DNA 序列将在第 8 章详细讨论。

着丝粒是 DNA 复制后染色体准确分离所必需的。复制后的染色体的两个拷贝叫姐妹染色体,它们必须分离为单拷贝,分别进入两个子细胞。与复制起始位点相似,着丝粒指导一个精细的蛋白质复合体的形成,此结构也称为动粒(kinetochore)。动粒的作用机制是与着丝粒 DNA 和与蛋白质纤维(微管)相互作用,拉动姐妹染色体相互远离并进入两个子细胞。与每条真核染色体上有多个复制起始位点不同,每条染色体上仅有一个着丝粒(图 7-7a)。若着丝粒缺失,复制后的染色体会随机分离,导致子细胞的染色体的缺失或加倍(图 7-7b)。如果一条染色体上存在一个以上拷贝的着丝粒,同样是灾难性的。动粒会附着在将染色体拉向相反方向的微管上,导致染色体的断裂(图 7-7c)。着丝粒的大小变化很大。在酿酒酵母中,着丝粒不到 200bp 为单一序列;而大多数真核生物的着丝粒大于 40kb,并且主要由重复 DNA 序列组成(图 7-8)。

端粒位于线性染色体的两端。与复制起始位点和着丝粒类似,端粒结合有许多蛋白质,在这种情况下,这些蛋白质执行着两个重要的功能。首先,端粒蛋白质可以将细胞中染色体的天然末端与染色体断裂和其他 DNA 的断裂区分开来。一般来说,DNA 的末端是经常发生重组和 DNA 降解的位点。结合在端粒上的蛋白质能避免重组和 DNA 降解的发生。其次,端粒作为一种特殊的复制起始位点,可以使细胞复制染色体的末端。

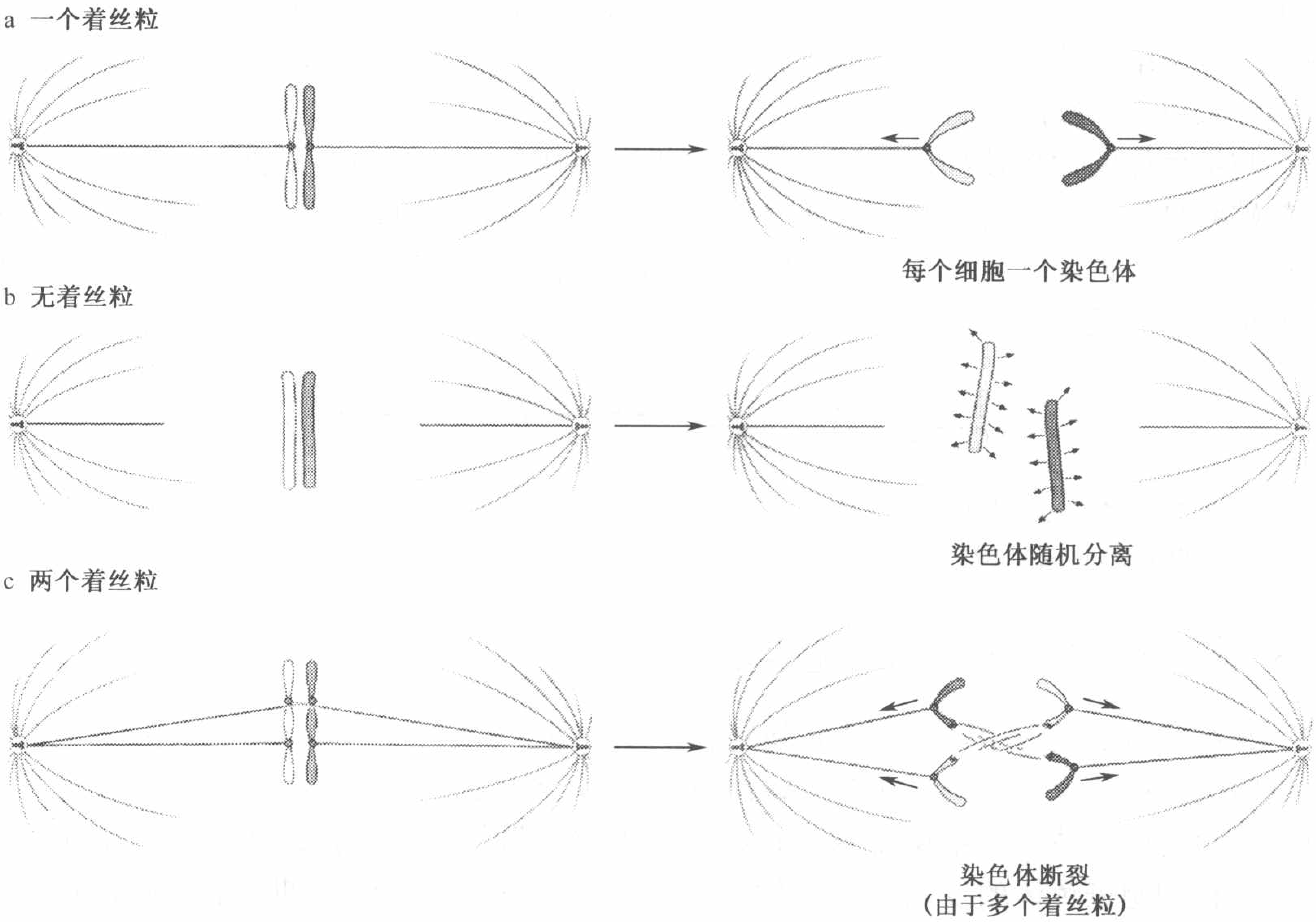


图 7-7 多于或少于一个着丝粒导致染色体的丢失或断裂。 a. 正常的染色体有一个着丝粒。染色体复制后，每个拷贝的端粒指导一个动粒的形成。这两个动粒结合到有丝分裂纺锤体的两极，并在细胞分裂之前向细胞相反的两边牵拉。b. 缺少着丝粒的染色体会很快从细胞中丢失。着丝粒缺失时染色体不能与纺锤体结合，只能随机地分配到两个子代细胞中。这经常导致一个子代细胞得到同一染色体的两个拷贝，而另一子代细胞丢失了这条染色体。c. 有两个或更多个着丝粒的染色体在分离时经常发生断裂。如果一个染色体有多个着丝粒，它能同时与纺锤体的两极结合。当分离起始时，有丝分裂纺锤体的两极的反向作用力经常会使同时结合在两极上的染色体断裂。

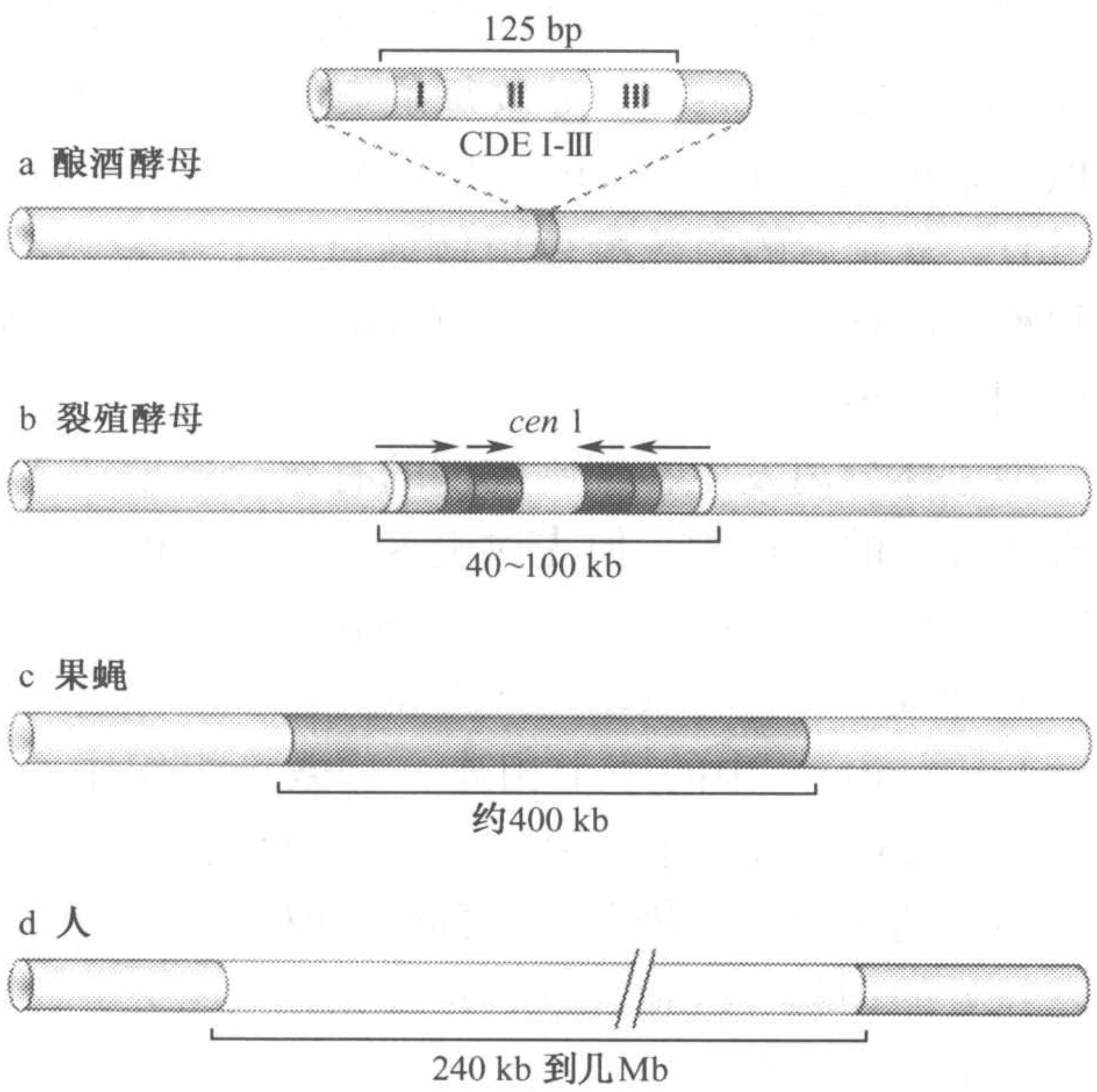


图 7-8 着丝粒的大小和组成在不同生物体间变化很大。 酿酒酵母的着丝粒很小且由非重复序列组成。而其他的生物体，如果蝇和裂殖酵母的着丝粒非常大而且主要由重复序列组成。裂殖酵母着丝粒仅中间的 4～7kb 是非重复序列，绝大多数果蝇和人的着丝粒由重复 DNA 组成。

在第 8 章里将讨论到标准的 DNA 复制机器不能完整地复制线性染色体的末端，原因将在第 8 章详细讨论。端粒通过一种特殊的 DNA 聚合酶——端粒酶（telomerase）来完成线性染色体末端的复制。

与染色体的大多数部分不同，端粒的相当一部分以单链的形式存在（图 7-9）。大多数端粒具简单的重复序列，重复序列因物种而异。典型的重复由一短的富 TG 的重复组成。例如，人的端粒具有 5′-TTAGGC-3′的重复序列。我们将在第 8 章中讲到，端粒中的重复序列是由于它们特殊的复制方式所造成的。



图 7-9 典型的端粒结构。框中显示了一个重复序列单元（取自人的细胞）。注意染色体 3′端的单链 DNA 区域可达数百个碱基长。

真核染色体的复制和分离发生在细胞周期的分裂期

在细胞分裂过程中，染色体必须复制并分离到子细胞中去。在细菌细胞中这些过程是同时发生的。也就是说，在 DNA 复制的同时，产生的两个拷贝会向细胞相反的两边分离。尽管已经清楚在细菌中这些过程受到严格的调控，但调控的机制还不清楚。相反，真核细胞染色体的复制和分离发生在细胞分裂过程的不同时期。下面我们将加以讨论。

细胞完成一轮分裂所需要的过程称为细胞周期（cell cycle）。大多数真核细胞的分裂会使子细胞维持与母细胞一致的染色体数目。这种分裂类型称为细胞的有丝分裂（mitotic cell division）。

有丝分裂的细胞周期可以分为 4 期：G_1、S、G_2 和 M（图 7-10）。涉及染色体增殖的关键过程发生在细胞周期的特定时期。DNA 合成发生在细胞周期的合成期或 S 期（synthesis 或 S phase），导致每条染色体的复制（图 7-11）。复制后配对的每条染色体叫做一条染色单体（chromatid），配对的两条染色单体称为姐妹染色单体（sister chromatid）。复制后的姐妹染色单体通过一个叫做黏粒（cohesin）的分子聚在一起，这一过程称为姐妹染色单体的附着（sister chromatid cohesion），这种状态一直维持到染色体的相互分离。

染色体分离发生在细胞周期的有丝分裂期或 M 期（mitosis，M phase）。下面将讲述有丝分裂的全过程，但先讨论这个过程中的 3 个关键步骤（图 7-12）。首先，每对姐妹染色单体与一个叫做有丝分裂纺锤体（mitotic spindle）的结构结合。纺锤体由长的蛋白质纤维——微管（microtubule）组成。微管与两个微管组织中心（microtubule organizing center）中的一个相连。[微管组织中心在动物细胞中称为中心粒（centrosome），在酵母和其他真菌中称作纺锤体极体（spindle pole body）。] 微管组织中心位于细胞的两侧形成“极”，微管向着两极拉动染色单体。染色单体的配对由组装在每个着丝粒处的动粒（kinetochore）所介导（图 7-6）。其次，由于黏粒的水解染色单体间黏附力的消失。在黏附力消失之前，它可以抵抗有丝分裂纺锤体的拉力。当黏附力消失之

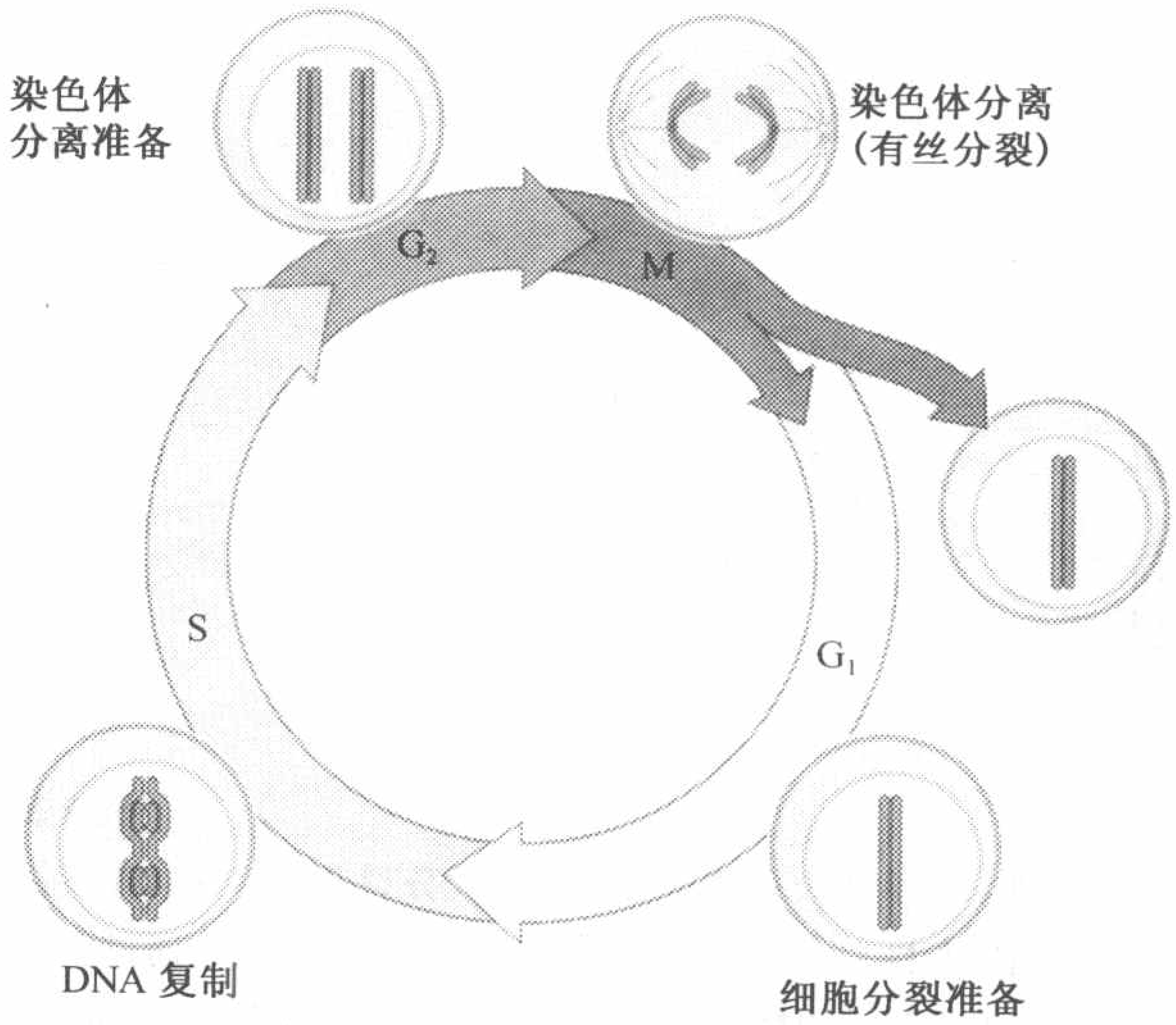


图 7-10 真核有丝分裂细胞周期。真核细胞周期有 4 个阶段。染色体复制发生在 S 期，染色体分离发生在 M 期。G_1 和 G_2 间期允许细胞为细胞周期的后续过程做准备。例如，许多真核细胞利用细胞周期的 G_1 期合成完成细胞分裂所需的足量的营养成分。

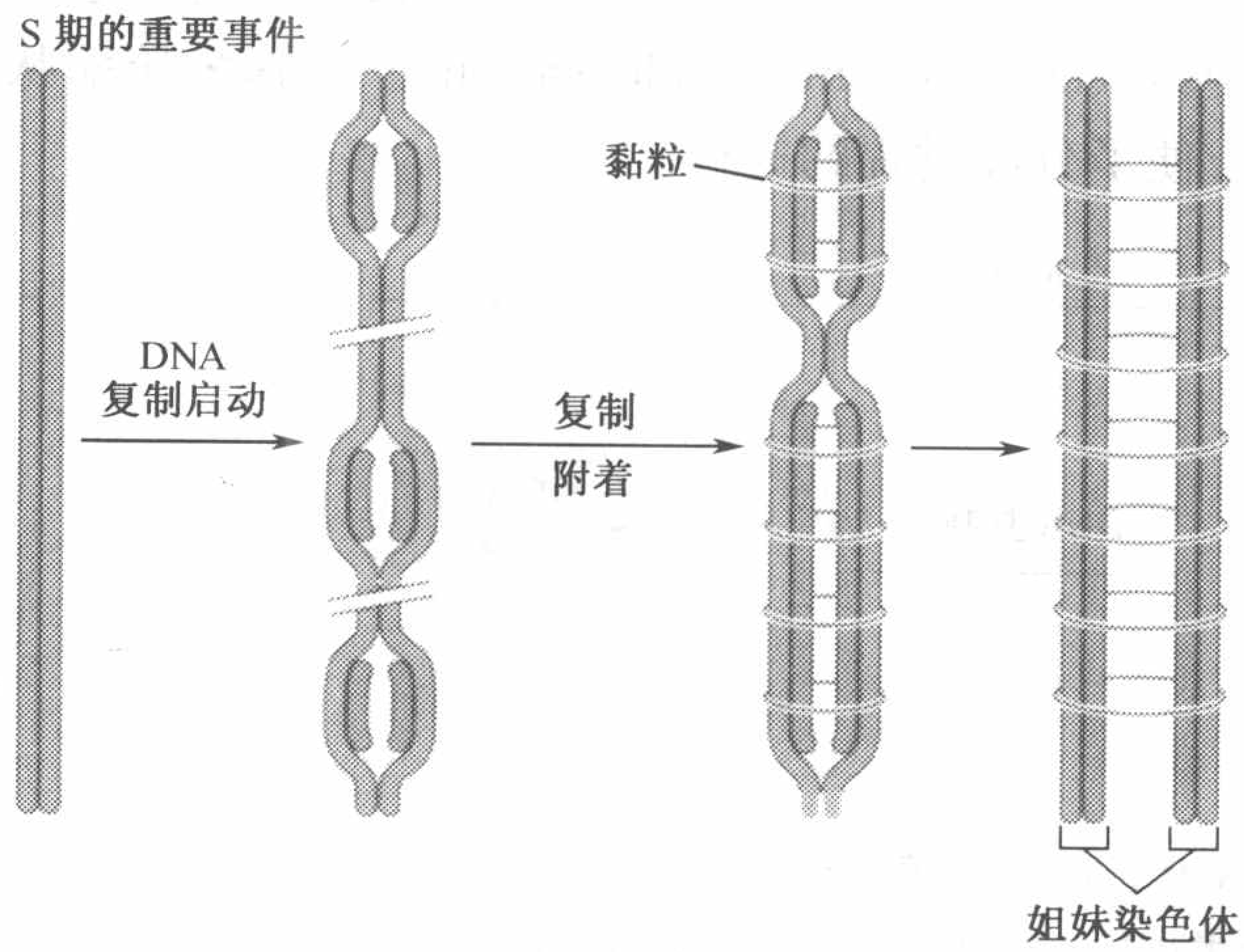


图 7-11 S 期的事件。在 S 期有两个主要的染色体事件发生。每条染色体完整地进行 DNA 复制拷贝，复制发生后不久，环状的黏粒分子将复制后的两拷贝 DNA 缚在一起，导致姐妹染色单体的附着。每个蓝色或红色的“管”代表一个单链 DNA 分子。

后，有丝分裂的第三个重要过程才能发生：姐妹染色单体分离（sister chromatid separation）。由于缺少了染色单体黏附力的抗衡，染色单体被迅速地拉向有丝分裂纺锤体相反的两极。因此，姐妹染色单体间的黏附力与有丝分裂纺锤体相反两极对姐妹染色单体动粒的黏附力是相互对立的作用，只有精细地协调二者才能使染色体正确地进行分离。

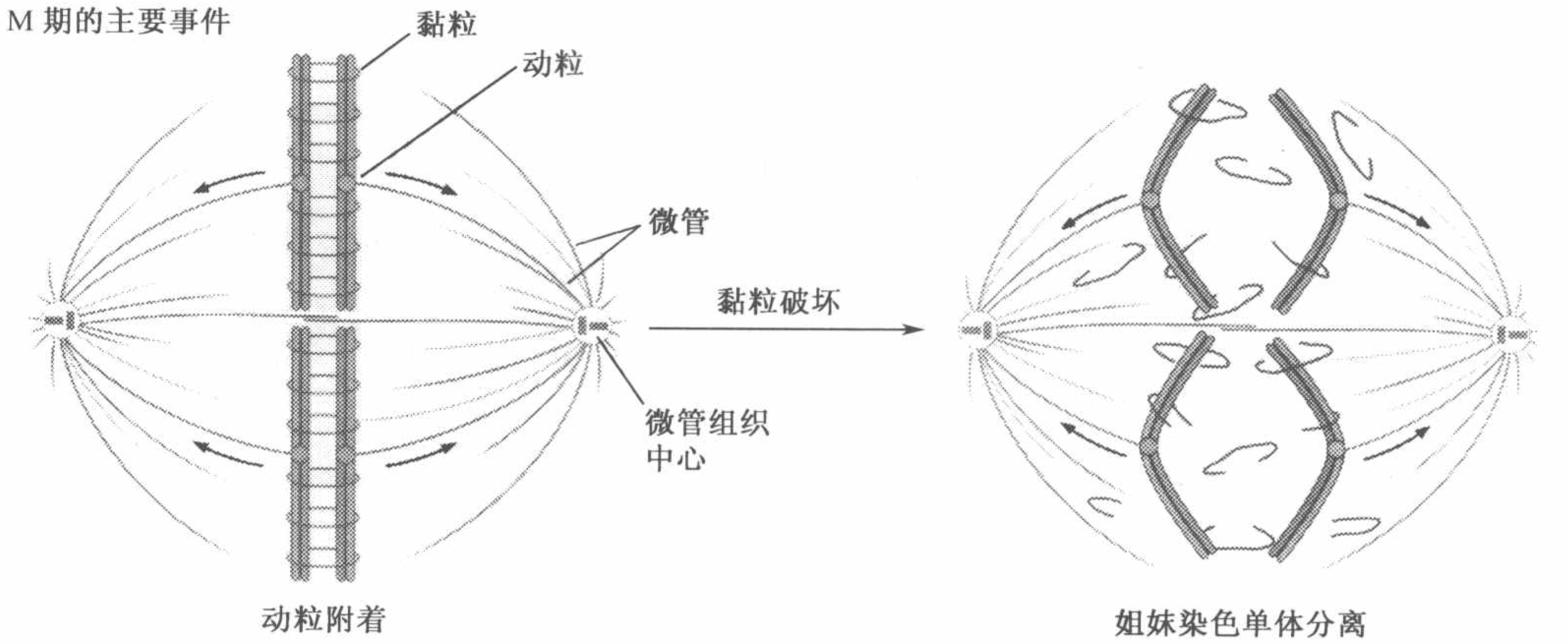


图 7-12 分裂期（M 期）的事件。三个主要的事件发生在有丝分裂期。首先，连接姐妹染色单体对的两个动粒分别黏附到有丝分裂纺锤体的相反两极。一旦全部的动粒都结合到两极上，通过破坏黏粒环使姐妹染色单体间的黏附力消失。最后，当黏附力消失后，姐妹染色体分离到有丝分裂纺锤体相反的两极。

真核细胞分裂时染色体的结构发生变化

在一轮细胞分裂中，染色体的结构发生多次变化；然而，染色体主要以两种状态存在（图 7-13）。当细胞的染色体分离时，染色体处于最压缩的状态。形成这种压缩状态的过程称为染色体凝聚（chromosome condensation）。在这种压缩状态下，染色体间完全相互独立，大大方便了分离过程的进行。

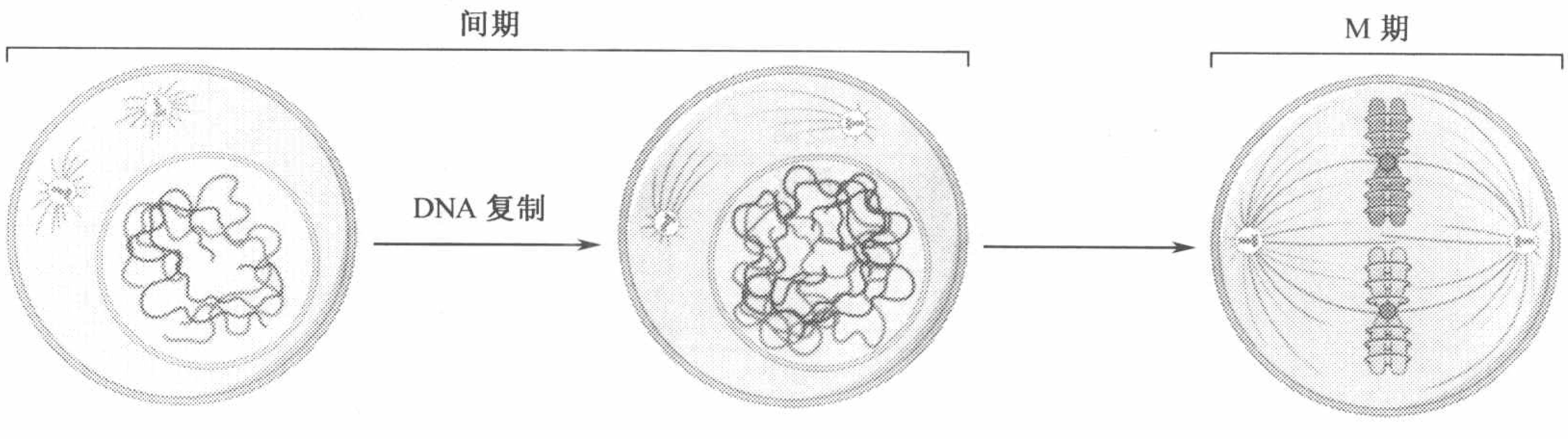


图 7-13 染色质结构的变化。M 期的染色体达到最大程度的压缩，而在细胞周期的其他阶段（有丝分裂的 G_1、S 和 G_2）染色体处于解聚状态。这些解聚的时期统称为间期。

在 G_1、S 和 G_2 期（总称为间期），染色体不分离时染色体的压缩程度大大降低。事实上，在细胞周期的这些阶段，染色体可能是高度缠绕的，与有丝分裂过程中染色体的有组织排布相比，此时的染色体更像一盘意大利面条。然而，在这些阶段中染色体的结构也会发生改变。DNA 复制要求与染色体结合的蛋白质几乎全部解离和重装。DNA 复制完成后，姐妹染色单体的黏附立即形成，使新复制的两条染色单体相互连接。在细胞周期中，单个基因转录的开关或者上下调伴随着染色体相关位置上结构的改变。因此，染色体是一个不断变化的结构，它更像是一个细胞器，而不只是一条简单的 DNA 链。

SMC 蛋白介导姐妹染色单体的黏附和染色体的凝聚

介导姐妹染色单体黏附和染色体凝聚的关键蛋白是相互关联的。染色体结构维持（SMC）蛋白是一类延伸蛋白（extended protein），它们通过长的螺旋卷曲结构域相互作用形成固定的配对（见图 5-15）。SMC 蛋白与非 SMC 蛋白一起形成多蛋白复合体，将两个 DNA 螺旋连接在一起。如前所讨论，一种含有 SMC 蛋白的复合体叫做黏粒（cohesin），DNA 复制后，黏粒将两个子代 DNA（姐妹染色单体）连接起来，这种连接是姐妹染色体黏附的基础。黏粒的结构被认为是一个由两个 SMC 蛋白和两个非 SMC 蛋白组成的大环。尽管姐妹染色体单体黏附的确切机制尚不清楚，已经建立了一个很好的模型，表明姐妹染色单体黏附的机制是两条姐妹染色单体穿过黏粒蛋白环的中心（图 7-14）。在这个模型中，黏粒的非 SMC 亚基发生蛋白水解会导致黏粒的开环和消失。

伴随染色体分离所发生的染色体凝聚同样需要一个相关的含有 SMC 蛋白的复合体——凝聚蛋白（condensin）。尽管我们对这种复合体的结构和功能知之甚少，但它具有许多与黏粒复合体相同的特点，因此猜测它也是一个环状复合体。如果是这样，它可能是利用环状的特点引起染色体的凝聚。例如，凝聚蛋白能将同一染色体的不同区域连接在一起，大大减少染色体的线性长度（图 7-14）。

有丝分裂维持亲本染色体的数目

我们现在讲述有丝分裂的全过程。有丝分裂分为几个时期（图 7-15）。在分裂前期（prophase），染色体凝聚成高度紧密的结构以准备进行分离。在前期结束时，核膜裂解，细胞进入分裂中期（metaphase）。

在分段中期，有丝分裂纺锤体形成，姐妹染色体的动粒与微管结合。只有当一对姐妹染色单体的两个动粒分别与从相对的微管组织中心发出的微管结合后，正确的染色单体连接才告完成。这种连接类型叫做二价联会（bivalent attachment，图 7-15），会导致微管对染色单体对施加拉张力，向相反的方向牵拉姐妹染色单体。从同一微管组织中心发出的微管与两个染色单体或者其中之一的结合，叫做单价联会（monovalent attachment）。这种结合不能产生拉张力，最终导致染色体的丢失。如果二价联会不能随后形成，单价联会能让染色体的两个拷贝拉到一个子细胞中去。姐妹染色单体的黏附力抵消了二价联会产生的拉张力，从而使所有染色体排列在细胞的中部，位于两个微管组织中心之间（这个位置被称为中期板，亦称赤道板）。值得一提的是，只有在所有的姐妹染色单体都进行二价联会后，染色体才会开始分离。

黏粒分子因蛋白质水解的破坏，导致姐妹染色单体间黏附的消失，从而触发染色体的分离。这一过程发生在细胞分裂的后期（anaphase）。在后期，姐妹染色单体分离并移动到细胞的两侧。一旦两个姐妹染色单体不再黏附在一起，它们便不能抵抗纺锤体微管向外的拉力。二价联会保证了每对姐妹染色单体对的两个成员被拉向相反的两极，使每个子代细胞都得到每个复制的染色体的一个拷贝。

有丝分裂的最后阶段称为末期（telophase）。在末期，包围每套分离的染色体的核膜重新形成。此时，两个细胞所共有的细胞质发生物理分离，细胞分裂完成。这个过程称为胞质分裂（cytokinesis）。

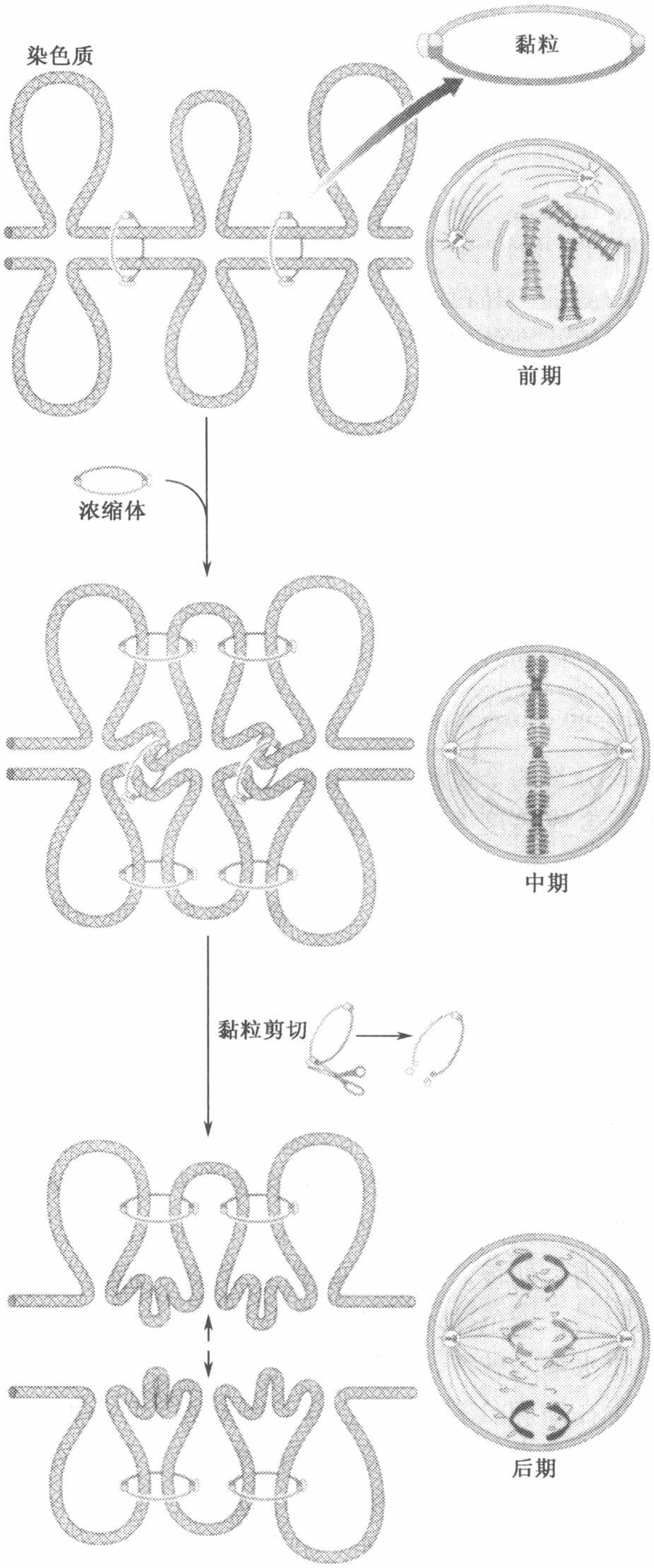


图 7-14 推测的黏粒和凝聚蛋白模型。黏粒和凝聚蛋白是环状蛋白复合体。包括两个 SMC 蛋白，细胞核骨架的组成部分。对连接远距离的或不同区域的 DNA 有重要作用。推测的含有环状结构的这些蛋白质允许 DNA 的两个区段柔韧且牢固的连接。在图示中，SMC 蛋白用绿色（黏粒）或者蓝色（凝聚蛋白）表示。（来源：Haering C. H. et al. 2002. Mol. Cell 9：773-778，F8，page 785，© Elsevier.）

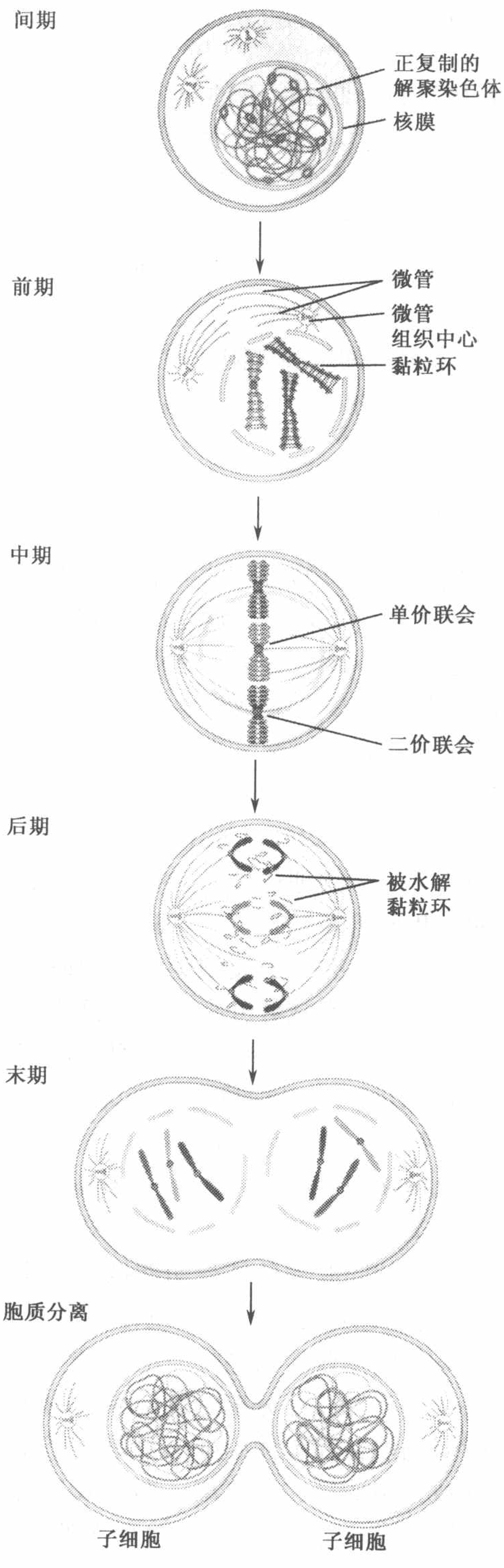


图 7-15　有丝分裂。在有丝分裂之前，染色体以解聚状态存在，此时为间期。前期，染色体凝聚，准备分离，在大多数真核细胞中，包围着染色体的核膜发生裂解；在中期，每个姐妹染色单体对结合到有丝分裂纺锤体的两极；进入后期，姐妹染色单体间黏附力消失，导致姐妹染色单体的分离；末期，染色体凝聚状态消失，包围着两群分离后的染色体的核膜重新形成；细胞质分裂是细胞周期的最后阶段，包围着两个核的细胞膜收缩，并最终分裂成两个子细胞。所有的 DNA 分子都是双链的。

细胞周期的间期为下一个细胞周期作准备，同时检查上一个阶段是否正确完成

真核细胞周期中余下的两个阶段是间期。G_1 期发生于 DNA 合成之前，G_2 期位于 S 和 M 期之间。细胞周期的间期有两个作用：一是为下一个细胞周期阶段作好时间上的准备，二是检查上一个细胞周期阶段是否正确完成。例如，在进入 S 期以前，大多数细胞必须达到一定的大小和蛋白质合成水平，以满足完成下一轮 DNA 合成所需的足够的蛋白质和营养成分。如果在细胞周期的前一阶段存在问题，细胞周期检查点（cell cycle checkpoint）则停止细胞周期并给细胞提供时间以完成这一步。例如，含有损伤 DNA 的细胞将细胞周期阻止在 DNA 合成之前的 G_1 期或者有丝分裂之前的 G_2 期，以防止产生损伤的染色体的事件发生。在细胞周期继续进行之前，这种延迟为损伤的修复提供了时间。

减数分裂减少了亲本染色体的数目

真核细胞分裂的另一种类型——减数分裂，专一地产生子代细胞，和母细胞相比，它仅含有半数的染色体。这是由 DNA 复制之后的两轮染色体分离完成的。与有丝分裂细胞周期相似，减数分裂细胞周期（meiotic cell cycle）包括 G_1 期、S 期和一个延长的 G_2 期（图 7-16）。在减数分裂的 S 期，每条染色体复制。与有丝分裂 S 期相同，姐妹染色单体仍然结合在一起。进入减数分裂的细胞必须是二倍体，因此 DNA 复制前每条染色体有两个拷贝，分别来自父母。DNA 复制之后，这些相关的姐妹染色单体对，称为同源染色体（homologs），互相配对并发生重组。同源染色体间的重组在两个同源染色体间构建了一个物理连接，这种连接是染色体分离时连接两个相关的姐妹染色单体对所必需的。我们将在第 10 章详细讨论减数分裂过程中的重组。

减数分裂与有丝分裂细胞周期的最大不同发生在染色体分离阶段。有丝分裂仅有一轮染色体分离，而减数分裂中的染色体经过两轮分离，称作减数分裂Ⅰ期和Ⅱ期。与有丝分裂相同，每一轮染色体分离都包括前期、中期和后期。减数分裂Ⅰ的中期也叫做中期Ⅰ(meiosis Ⅰ)，此时同源染色体与微管纺锤体相反的两极结合。这一结合由动粒介导。由于每个姐妹染色单体对的两个动粒都结合到微管纺锤体的同一极，这个相互作用被称为单价联会（monovalent attachment）（与有丝分裂的二价联会相反，有丝分裂中每个姐妹染色单体对的两个动粒是结合到纺锤体相反的两极上的）。如同在有丝分裂中一样，配对的同源染色体最初能抵抗纺锤体所施加的将它们拉开的拉张力。在减数分裂Ⅰ期时，此过程是通过同源染色体间的物理连接或者重组所诱发的染色体交换所介导。沿姐妹染色单体臂上的黏附力也可提供这种抵抗力。当这种黏附力在后期Ⅰ消失后，同源染色体相互释放并分离到细胞相反的两极。重要的是，姐妹染色单体的着丝粒附近间的黏附力仍然存在，使姐妹染色单体仍保持配对状态。

减数分裂的第二轮分离，即减数分裂Ⅱ期（meiosis Ⅱ），与有丝分裂非常相似，主要的不同是这轮分离之前没有 DNA 的复制。相反，纺锤体已形成且与两个新分离的姐妹染色单体对相连。与有丝分裂相同，中期Ⅱ（metaphase Ⅱ）时的纺锤丝以二价联会方式结合到姐妹染色单体对的两个动粒上。在减数分裂Ⅰ后仍维持在着丝粒附近的黏附

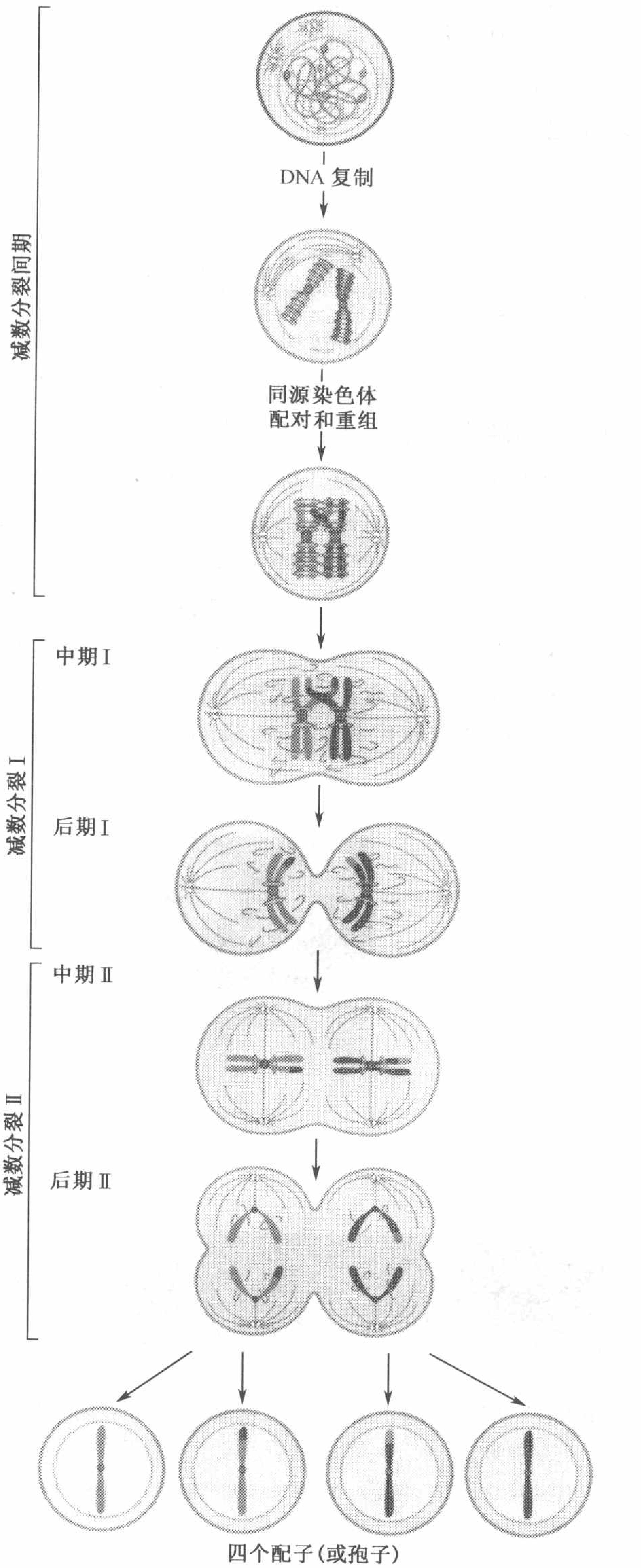


图 7-16 减数分裂。减数分裂与有丝分裂相似，也能分成不连续的阶段。DNA 复制后，同源姐妹染色单体相互配对，形成 4 条相关的染色体。为简便起见，仅用一个染色体描述分离过程，蓝色和黄色拷贝代表分别来自父母。在配对过程中，来自不同姐妹染色质的染色单体发生重组，在同源染色体间形成一种称为交叉的连接。中期Ⅰ期间，每个姐妹染色单体对的两个动粒与减数分裂纺锤体的一极结合。同源的姐妹染色单体的动粒与相反的两极结合形成的拉张力，被同源染色体间的连接力所抵抗。进入后期Ⅰ，两个过程共同导致同源染色体间的相互分离，即作用于染色体臂的染色单体黏附力的消失和同源染色体间交叉的解离。共同作用使两者同源染色体相互分离。姐妹染色单体通过着丝粒附近的黏附力仍然结合在一起。减数分裂Ⅱ期与有丝分裂非常相似。在减数中期Ⅱ，两个减数分裂纺锤体形成。与有丝分裂中期相同，每个姐妹染色单体对的动粒与减数分裂纺锤体的相反两极结合。在后期Ⅱ，姐妹染色单体间残余的黏附力消失，姐妹染色单体相互分离。分离后的 4 套染色体包装到细胞核内并分离成 4 个细胞，形成 4 个孢子或配子。所有的 DNA 分子以双链形式存在。(来源：Murray A. and Hunt T. 1993. The cell cycle: The introduction, Fig. 10.2 Copyright© 1993 by Oxford University Press, Inc. 经 Oxford University Press, Inc. 许可)

力是抵抗纺锤体拉力的关键因素。当这种黏附力消除后，后期Ⅱ（anaphase Ⅱ）开始。此时的细胞中有 4 套染色体，每套只包含每条染色体的一个拷贝。每套染色体外周形成一个细胞核，然后胞质分离，形成 4 个单倍体细胞。这些细胞可通过交配形成新的二倍体细胞。

在显微镜下可观察到不同时期的染色体结构

长期以来，显微镜就用来观察染色体的结构和功能。事实上，在了解染色体是细胞中遗传信息的载体很久之前，对细胞分裂过程中染色体的运动和变化就有了较深入的研究。用简单的光学显微镜就可以容易地观察到凝聚状态下的有丝分裂或减数分裂染色体。凝聚染色体的显微镜分析还常被用于鉴定人类细胞的染色体构成，以检测是否存在如染色体缺失等染色体异常，或者某一个染色体的拷贝过少或过多的个体。

处于非有丝分裂期（即间期）的染色体 DNA 凝聚程度较低（图 7-17a）。在电子显微镜下可以观察到两种状态的染色质：直径为 30nm 的纤丝和直径为 10nm 的纤丝（图 7-17b）。30nm 的纤丝是染色质较紧密的结构，通常折叠成大环从蛋白核或骨架中伸出；相反，10nm 的纤丝是染色质较松散的形式，通常组装成一串串有规则的“念珠”，而这些珠子就是核小体这些蛋白质-DNA 结构体在染色体结构和功能的调控。我们首先着重讲述核小体的特点，包括核小体是怎样形成的，然后描述依赖于核小体的结构是怎样对核 DNA 的易接近性进行控制的。

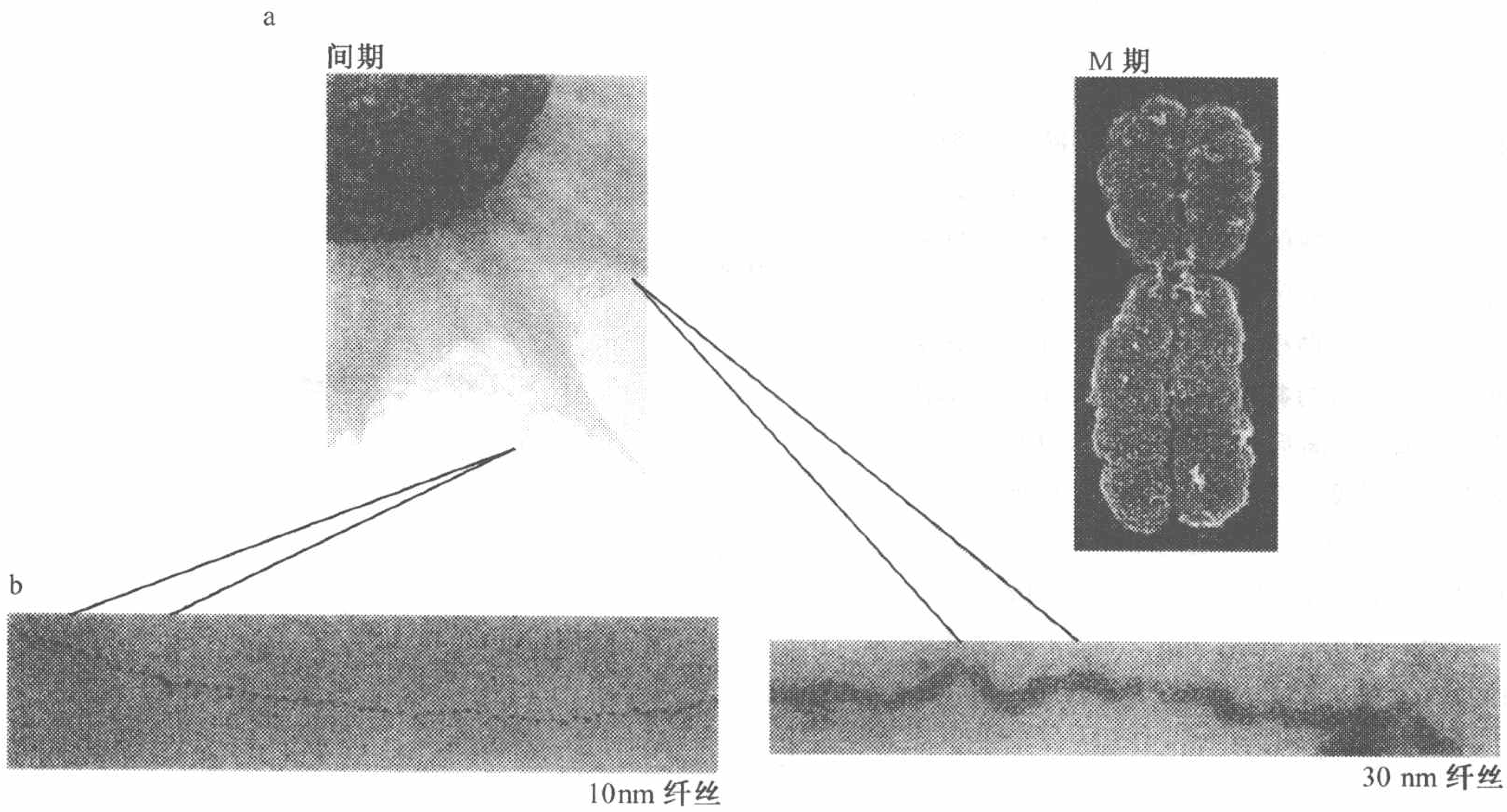


图 7-17 电子显微镜下观察到的染色质结构。 a. M 期和间期 DNA 的电子显微照片显示了染色质的结构变化。b. 间期细胞中不同形式的染色质的电子显微照片显示了 30nm 和 10nm 的染色质纤丝（念珠状）。（来源：a. Alberts B. et al. 2002. Molecular biology of the cell, 4th edition. Figs 4-21和4-23, Garland Science/Taylor & Francis LLC@V. Foe.）

核小体

核小体是染色体的结构单位

在真核细胞中大多数DNA被包装、核小体。核小体由8个组蛋白所形成的核心组成，DNA缠绕在组蛋白核心上。每个核小体之间的DNA（“念珠”上的“线”）叫做连接DNA（linker DNA）。DNA通过包装进核小体，被压缩了大约6倍，但这远不及在真核细胞中观察到的压缩1000～10 000倍的DNA。然而，这种DNA的初始包装对于后续各个层次上的DNA压缩是至关重要的。

与核小体结合最紧密的DNA叫做核心DNA（core DNA），它像线缠绕线轴一样盘绕组蛋白八聚体约1.65圈（图7-18）。通过核酸酶处理能确定每个核小体上DNA的长度（框7-1）。在所有的真核细胞中，长度约147bp的核心DNA是核小体一个不变的特征。相反，核小体之间的连接DNA的长度是可变的。这个距离一般是20～60bp，并且每个真核生物有各自特征性的连接DNA的平均长度（表7-4）。连接DNA平均长度的不同可能是因为不同生物由核小体DNA形成的更复杂结构的特点不同，而不仅仅是核小体自身的不同（参见染色体高级结构部分）。

在任何细胞中都会有数段DNA没有被包装到核小体中。这些DNA区段一般参与基因的表达、复制或者重组。这些位点尽管未与核小体结合，但通常参与调控或参与这些过程的非组蛋白结合。我们将在随后的第17章讨论从DNA上去除核小体并维持此DNA区域在无核小体状态下的机制。

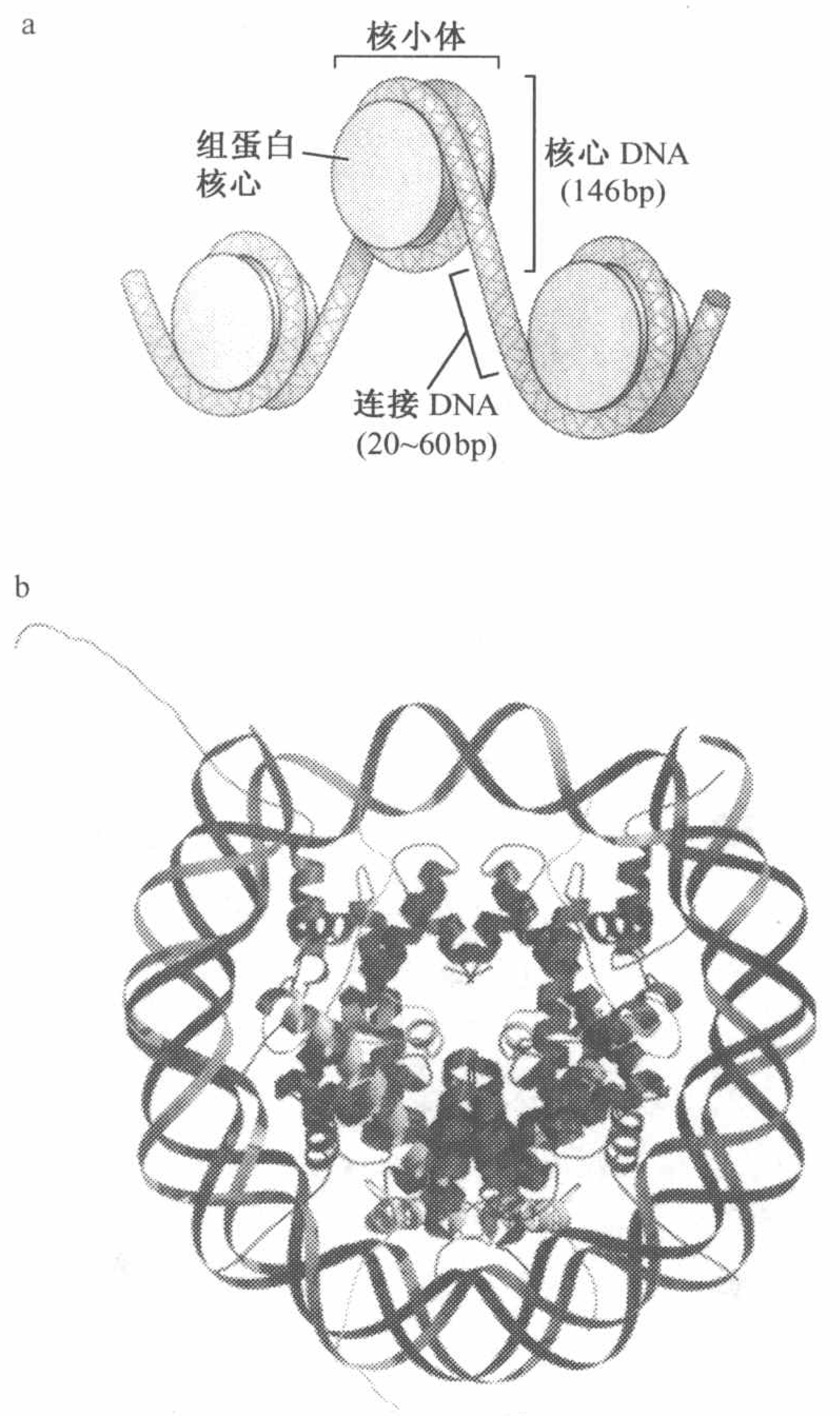
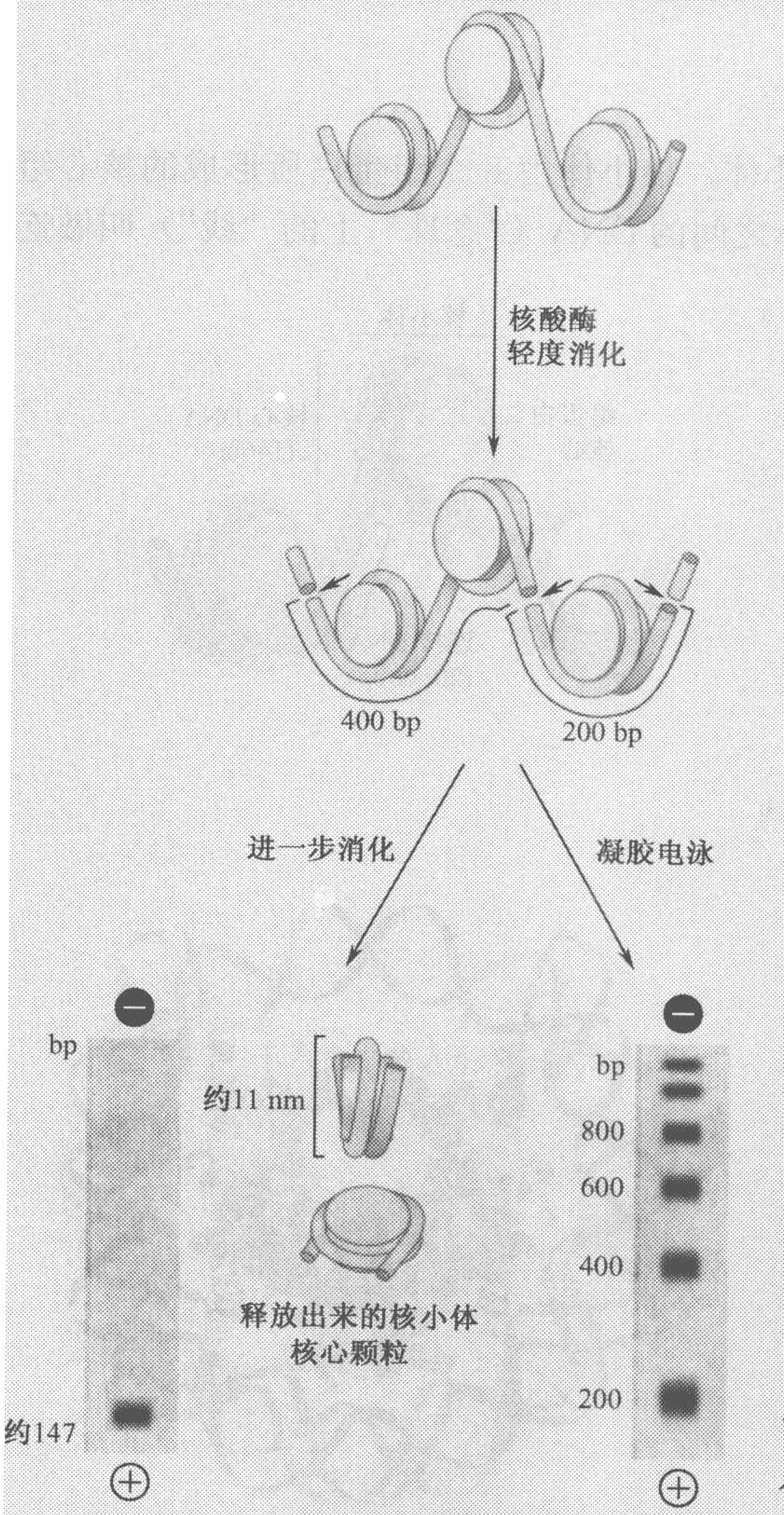


图7-18 DNA包装成核小体。a. 核小体包装和组织的示意图。b. 一个核小体的晶体结构显示了DNA盘绕在组蛋白核上。H2A、H2B、H3和H4分别用红色、黄色、紫色和绿色表示。注意在随后的图中不同的组蛋白都用相同的颜色表示。（Luger K. et al. 1997. Nature 389：251-260。）镜像图用MolScript和Raster3D制作。

框 7-1　微球菌核酸酶和核小体 DNA

通过利用无序列特异性的微球菌核酸酶（micrococcal nuclease，MNase）处理染色体，人们第一次得到了纯化的核小体。这个酶剪切 DNA 的能力主要由 DNA 的易接近性控制。微球菌核酸酶能迅速地剪切无蛋白质保护的 DNA 序列而对结合有蛋白质的 DNA 序列的剪切效率很差。利用这种酶有控制地处理染色体，可得到具核酸酶抗性的、与组蛋白结合的 DNA 分子。这些 DNA 分子的长度为 160～220bp，且与各两个拷贝的 H2A、H2B、H3 和 H4 结合。一般来说，这些颗粒包括与核小体紧密结合的 DNA 和一单位的连接 DNA。进一步的微球菌核酸酶处理会使全部的连接 DNA 降解，剩下的最小的核小体仅包括 147bp 的 DNA，被称为核小体核心颗粒（nucleosome core particle）。

利用一个简单的实验可以测量缠绕每个核小体的 DNA 的平均长度（框 7-1 图 1）。用微球菌核酸酶温和地处理染色质，这将在部分的连接 DNA 仅仅打开一个切点。核酸酶处理之后，从全部的蛋白质（包括组蛋白）中提取 DNA 并进行凝胶电泳分离。电泳得到一组具长度梯度的片段，其长度为核小体到核小体的平均距离的整倍数。这是由于微球菌核酸酶对染色质的部分消化所造成的。因此，有许多核小体未被消化分离，使得其 DNA 片段等于结合于这些核小体上的全部 DNA 的长度。进一步的消化将切开全部的连接 DNA，形成核小体核心颗粒，这时仅得到一条约 147bp 的片段。



框 7-1 图 1　微球菌核酸酶对核小体 DNA 的逐步消化。（R. D. Kornberg 惠赠）

表 7-4　各种生物体中连接 DNA 的平均长度

物种	核小体重复长度/bp	连接 DNA 的平均长度/bp
酵母	160～165	13～18
海胆（精子）	约 260	约 110
果蝇	约 180	约 33
人类	185～200	38～53

任何细胞中都含有没有包装成的长 DNA，一般来说，它们参与基因表达、复制或重组等功能。尽管这些长 DNA 没有核小体的结合，但这些结合位置在执行这些功能时通常都与非组蛋白相关，这些非组蛋白要么调控要么参与这些功能。我们将在第 17 章中讨论核小体从 DNA 上的移除机制以及这些 DNA 是如何保持无核小体状态的。

组蛋白是带正电荷的小分子蛋白质

组蛋白是目前所知与真核 DNA 相关的丰度最高的蛋白质。真核细胞一般包括 5 种组蛋白：H1、H2A、H2B、H3 和 H4。组蛋白 H2A、H2B、H3 和 H4 是核心组蛋白（core histone）两个拷贝的这几种组蛋白形成核小体的蛋白核，核小体 DNA 盘绕其上。组蛋白 H1 不是核小体核心颗粒的一部分。相反，它与连接 DNA 结合，被称为连接组蛋白（linker histone）。4 种核心组蛋白在细胞中以同等的数量存在，而 H1 仅有其他组蛋白丰度的一半。这与每个核小体（含每个核心组蛋白的两个拷贝）与 H1 一一对应的发现相符。

组蛋白与带负电荷的 DNA 分子紧密结合。与此一致的是，组蛋白中带正电荷的氨基酸的含量很高（表 7-5）。各种组蛋白中有超过 20%的氨基酸残基是赖氨酸或精氨酸。核心组蛋白是相对较小的蛋白质，分子质量一般为 11～15kDa，组蛋白 H1 稍大，分子量约为 20kDa。

表 7-5　组蛋白的一般特征

组蛋白类型	组蛋白	分子质量/Da	赖氨酸及精氨酸的百分含量/%
核心组蛋白	H2A	14 000	20
	H2B	13 900	22
	H3	15 400	23
	H4	11 400	24
连接组蛋白	H1	20 800	32

核小体的蛋白核是一个圆盘状的结构，只有当 DNA 存在时才组装成一种有序的结构。没有 DNA 时，核心组蛋白在溶液中形成中间组装体。在每种核心组蛋白中都存在一个保守区域，称为组蛋白折叠域（histone-fold domain），调节这些组蛋白中间体的组装（图 7-19）。组蛋白折叠由 3 个 α 螺旋组成，螺旋间由两个短的无规则的环所隔开。组蛋白折叠调节头尾相连的专一配对的组蛋白异源二聚体的形成。组蛋白 H3 和 H4 首先形成异源二聚体，然后两个二聚体形成一个四聚体，含有两个分子的 H3 和 H4。相反，H2A 和 H2B 在溶液中形成异源二聚体而不是四聚体。

一个核小体的组装涉及这些结构单位与 DNA 有序地结合（图 7-20）。首先，H3 · H4 四聚体与 DNA 结合；然后两个 H2A · H2B 二聚体结合到 H3 · H4-DNA 复合体，形成最后的核小体（图 7-18）。我们将在本章的后面部分讨论组装过程在什么时候、如何在细胞中完成的。

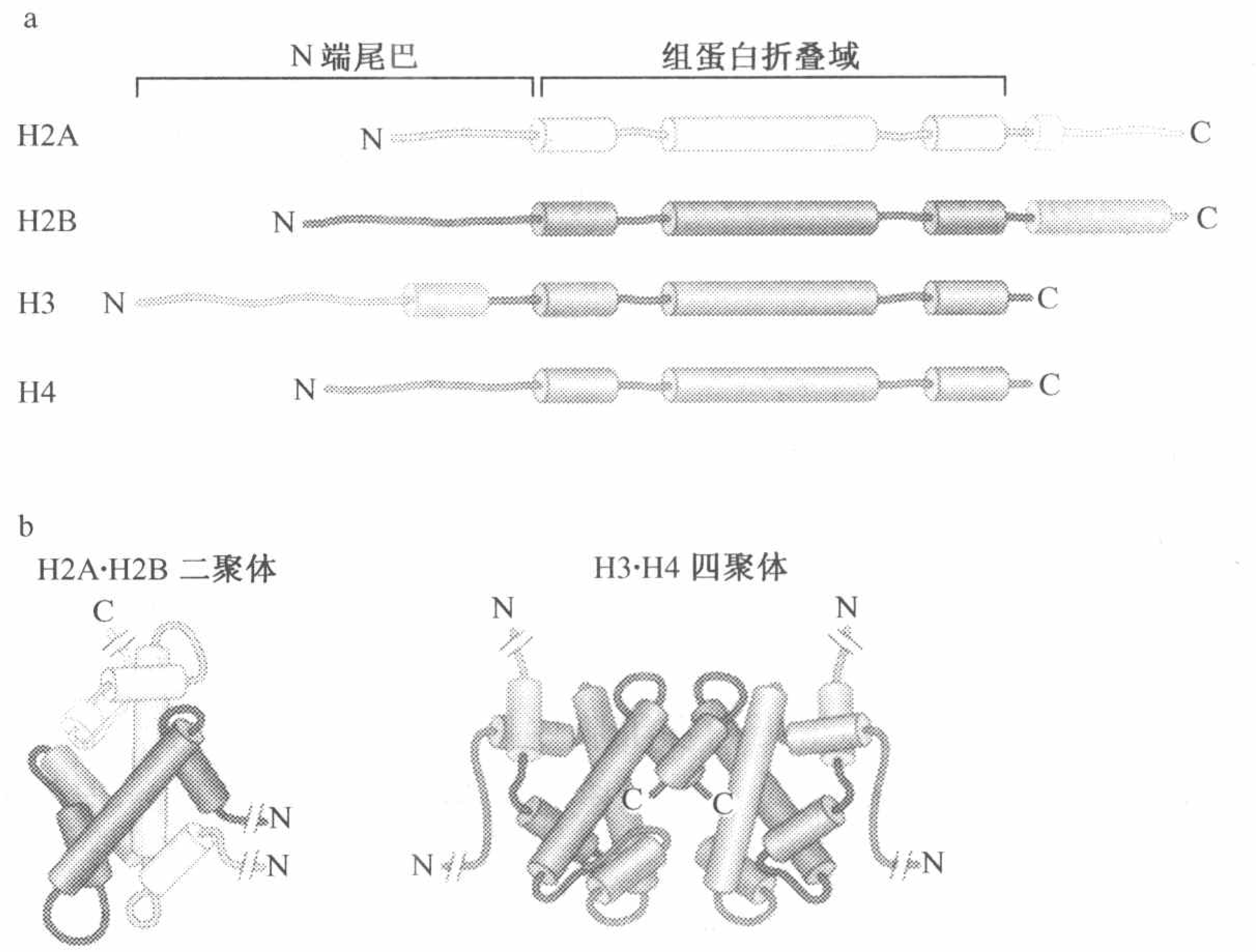


图 7-19　核心组蛋白具有共同的折叠结构。 a. 图中 4 种组蛋白显示为线性分子。形成 α 螺旋的组蛋白折叠域以圆柱表示，注意每个组蛋白都有结构上特异的相邻区域，其中包括其他的 α 螺旋区。b. 两个组蛋白（H2A 和 H2B）的螺旋区结合形成一个二聚体。H3 和 H4 也用相似的方式形成 $H3_2$ • $H4_2$ 四聚体。（Alberts B. et al. 2002. Molecular biology of the cell，4th edition，p. 209，Fig. 4-26. ©Garland Science Taylor & Francis. LLC.）

每个核心组蛋白有一个 N 端延伸，称为“尾巴”，这是因为它没有一个确定的结构，而且是完整的核小体中易接近的部分。用胰蛋白酶（能专一性地从蛋白质上去除带正电荷的氨基酸）处理核小体能检测到这种易接近性。用胰蛋白酶处理核小体能迅速地去除组蛋白易接近的 N 端尾巴，但不能切断紧密包装的组蛋白折叠区（图 7-21）。因为蛋白酶消化后 DNA 仍然与核小体紧密结合，所以暴露的 N 端尾巴对于 DNA 与组蛋白八聚体的结合是不必要的。然而，尾巴上有许多高度修饰化的位点，可以改变这些核小体的功能。这些修饰包括丝氨酸、赖氨酸和精氨酸残基上的磷酸化、乙酰化和甲基化。我们将在讨论核小体功能之后，再讨论组蛋白尾修饰的作用。现在，我们讨论核小体的详细结构。

核小体的原子结构

核小体核心颗粒（图 7-18b，147bp 的 DNA 加上一个完整的组蛋白八聚体）高分辨率的三维结构为我们揭示了核小体是如何行使功能的。核小体对 DNA 的高度亲和性、DNA 结合到核小体上时发生的扭曲和 DNA 序列的非特异性这些现象，都可用此组蛋白与 DNA 相互作用的结构来解释。这个结构也凸显了 N 端尾巴的功能和位置。最后，此 DNA 与组蛋白八聚体相互作用的结构揭示了核小体的动态特征和核小体组装的过程。在接下来的几部分中我们将讨论核小体的每一个特性。

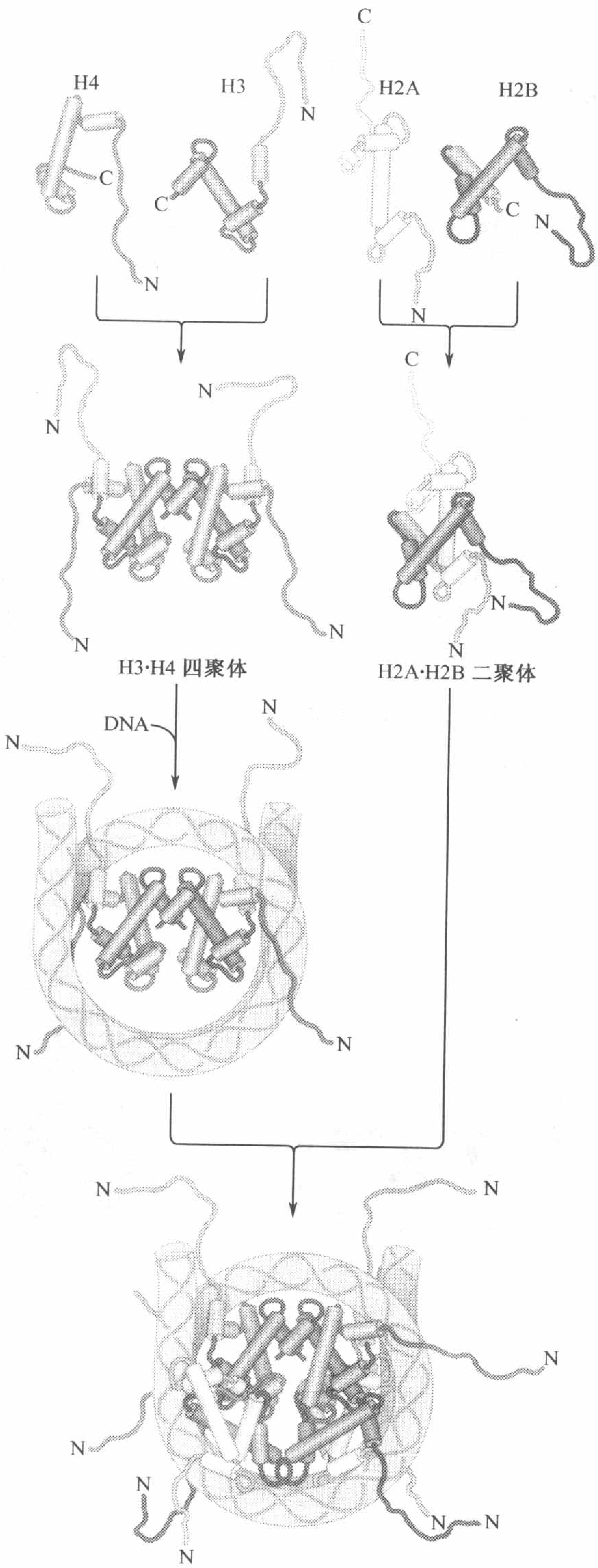


图 7-20　核小体的组装。$H3_2$ · $H4_2$ 四聚体的形成启动核小体的组装，四聚体然后与双链 DNA 结合。结合到 DNA 上的 $H3_2$，$H4_2$ 四聚体再与两个拷贝的 H2A · H2B 二聚体结合，完成核小体的组装。（来源：Alberts B. et al. 2002. Molecular biology of the cell，4th edition，p. 210，Fig. 4-27. Garland Science/Taylor & Francis LLC@J. Waterborg.）

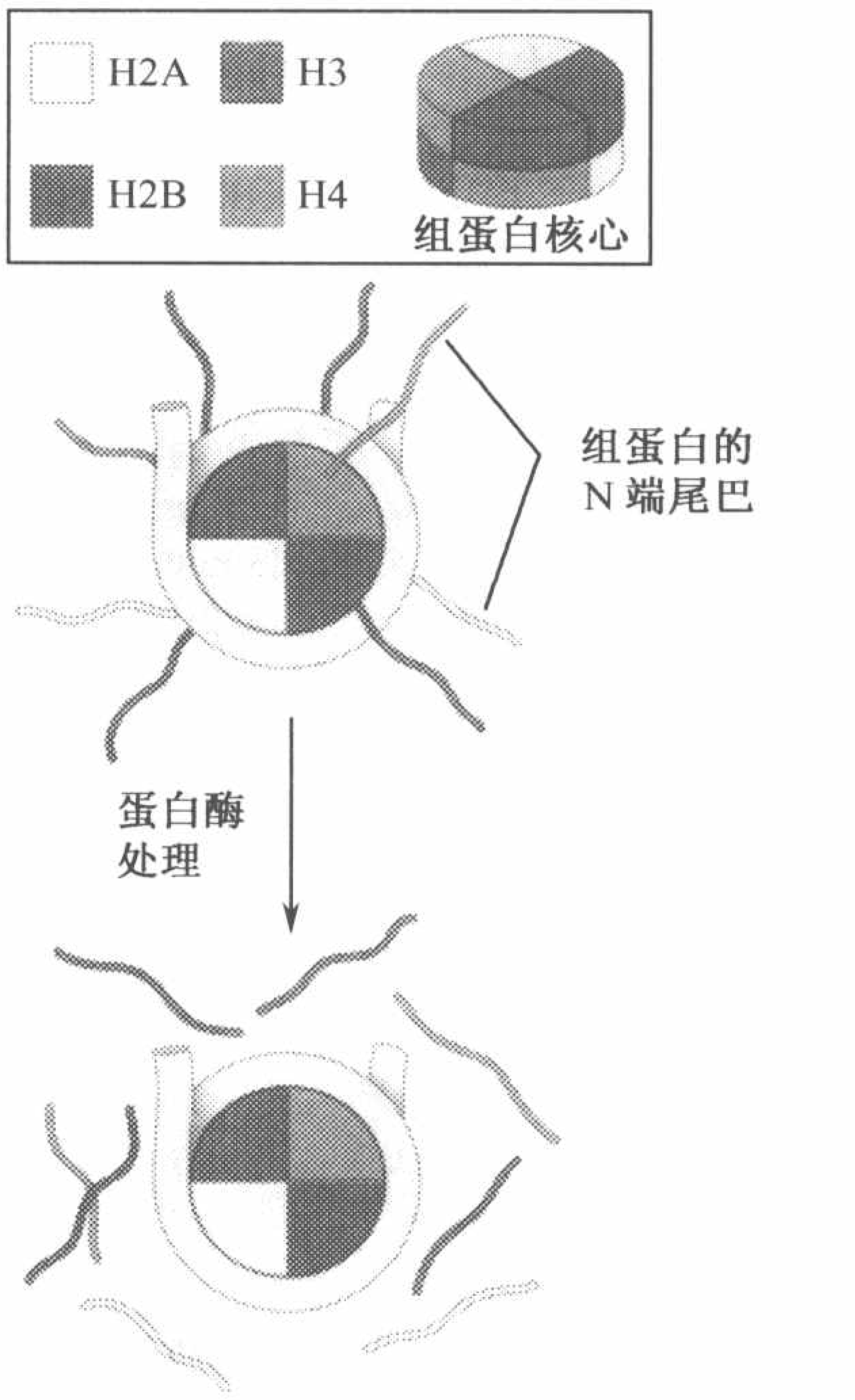


图 7-21 核心组蛋白的 N 端尾易受到蛋白酶的作用。用有限的可以在碱性氨基酸之后切开的蛋白酶（如胰蛋白酶）处理核小体，能专一地去除 N 端尾巴而保留完整的组蛋白核。

核小体中 DNA 的组蛋白结合区

尽管核小体不是完全对称的，但是有大致的二重轴对称，叫做二价轴（dyad axis）。如果将圆盘状的组蛋白八聚体的表面视作时钟，将 147bp 的 DNA 的中点定在 12 点钟的位置（图 7-22），就可以看到这种对称性，而 DNA 的末端靠近于 11 点和 1 点钟的位置。从 12 点到 6 点通过圆盘的中心画一条线，定义为二价轴。绕着这个轴将核小体旋转 180°，可以看到一个和旋转前几乎相同的外观。（见结构指南 7-1）

H3 · H4 四聚体和 H2A · H2B 二聚体分别与核小体中特定的 DNA 区段相互作用（图 7-23）。H3 · H4 四聚体的组蛋白折叠区与 147bp DNA 中间的 60bp 相互作用。H3 的 N 端区域最接近组蛋白折叠区，形成第 4 个 α 螺旋，与结合的 DNA 两个末端的 13bp 片段相互作用（这个区域与前述的无结构的 H3N 端尾是有区别的）。如果我们将核小体绘成上述的时钟，H3 · H4 四聚体构成组蛋白八聚体的上半部分。重要的是，组蛋白 H3 · H4 四聚体在核小体中占据着关键的位置，分别与

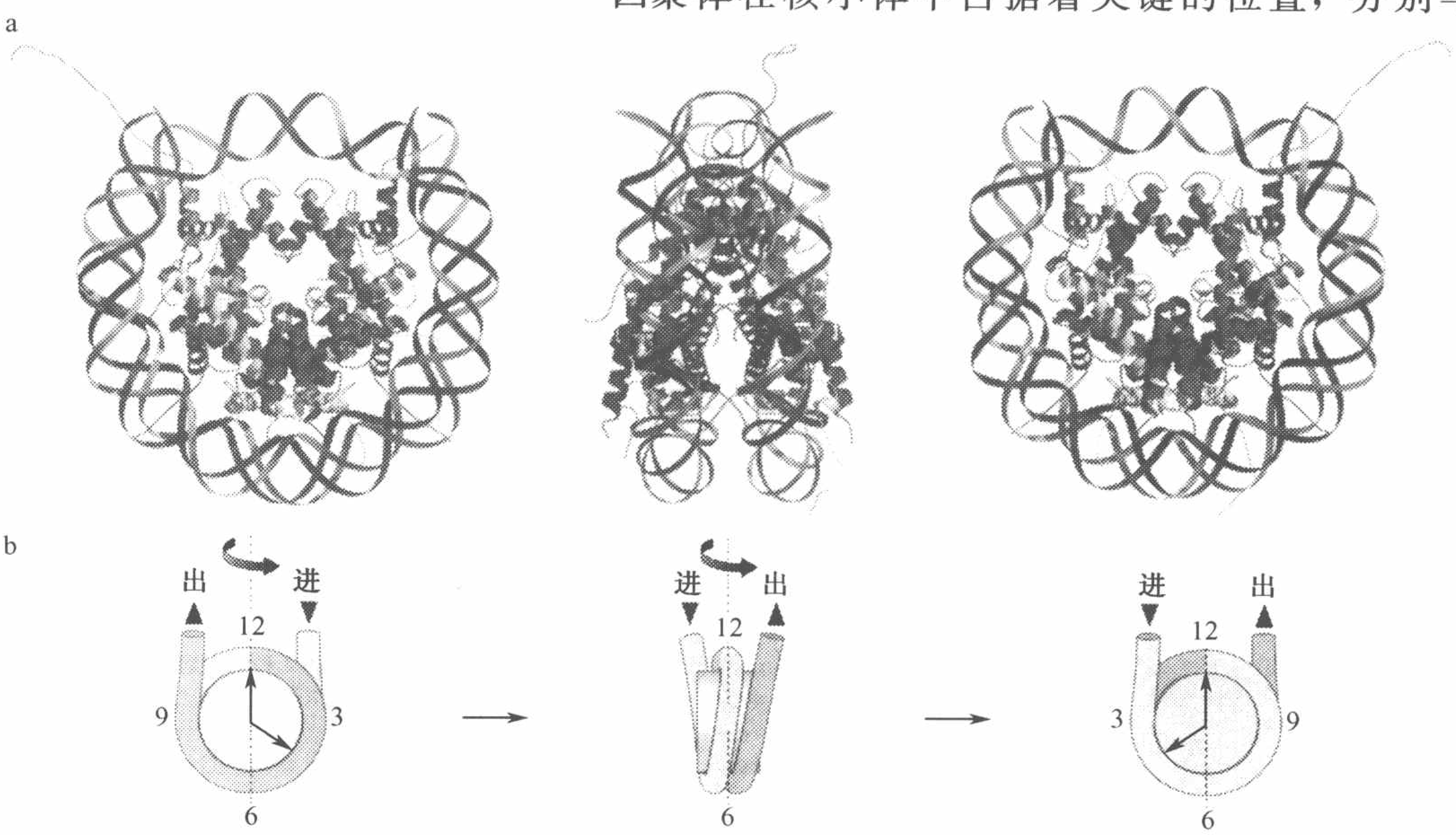


图 7-22 核小体有一个大致的二重轴对称。a. 三维结构；b. 核小体"钟面"示意图。从 3 个方向显示核小体的结构。每个方向都是绕图 7-22b 中所示的 12 点到 6 点的轴线旋转 90°所得。（a. Luger K. , et al. 1997. Nature 389:251-260.）镜像图用 BobScript、MolScript 和 Raster 3D 制作。

DNA 的中部和两端结合（图 7-23a 的标为青绿色的 DNA）。两个 H2A · H2B 二聚体分别与结合着 H3 和 H4 的中间 60bp 两侧大约各 30bp 的 DNA 结合。我们再次用时钟作比喻，与 H2A · H2B 二聚体结合的 DNA 定位于核小体圆盘的大约 5～9 点钟的位置。两个 H2A · H2B 二聚体构成组蛋白八聚体的底端部分，从 DNA 的末端穿过圆盘（图 7-23b 中标为橙色的 DNA）。

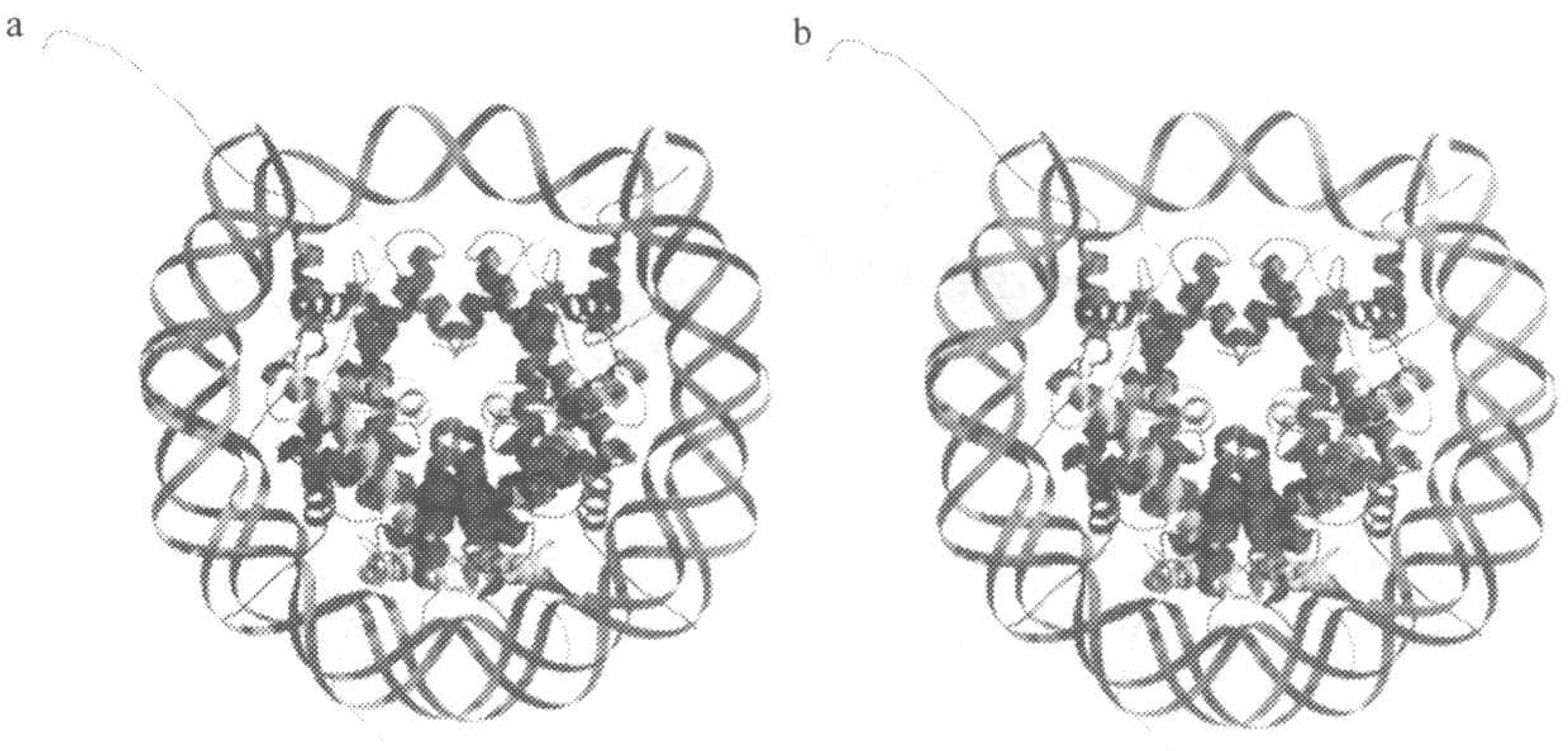


图 7-23　组蛋白和核小体 DNA 的相互作用。 a. H3 · H4 与 DNA 的中部和末端结合（用青绿色表示）。b. H2A · H2B 二聚体与核小体一侧 DNA 的 30bp 结合。与 H2A · H2B 二聚体结合的 DNA 用橙色表示。（Luger K. , et al. 1997. Nature 389:251-260. ）镜像图用 BobScript、MolScript 和 Raster 3D 制作。

H3 · H4 四聚体与 DNA 之间广泛的相互作用有助于解释核小体的有序组装（图 7-24）。H3 · H4 四聚体与 DNA 的中部和末端结合，造成 DNA 高度弯曲和固定，使得 DNA 与 H2A · H2B 二聚体的结合相对容易。与此相反，DNA 与 H2A · H2B 二聚体相对较短的结合长度，不足以造成 DNA 与 H3 · H4 四聚体的结合。

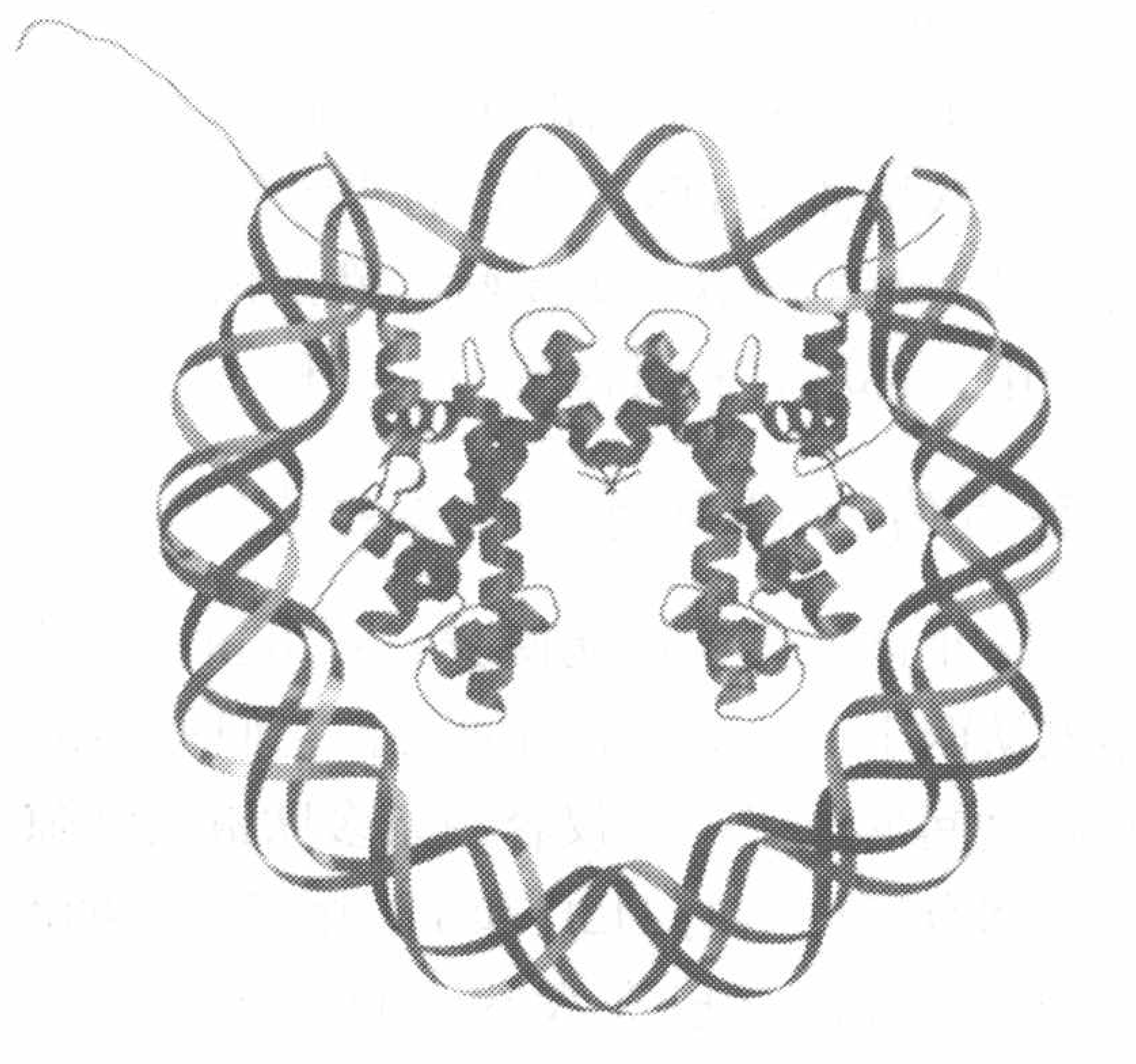


图 7-24　缺少 H2A 和 H2B 的核小体。 在这个图中组蛋白 H2A 和 H2B 被人为地从核小体中去除。这个结构可能与核小体组装过程中的 DNA · H3$_2$ · H4$_2$ 四聚体中间物类似（图7-20）。（Luger K. , et al. 1997. Nature 389:251-260. ）镜像图用 BobScript、MolScript 和 Raster 3D 制作。

许多不依赖于 DNA 序列的接触介导核心组蛋白与 DNA 间的相互作用

对组蛋白和核小体 DNA 相互作用的进一步观察揭示出核小体内的 DNA 结合和弯曲的结构基础。每次 DNA 的小沟面对组蛋白八聚体时可发现一个接触位点，共发现了 14 个不同的位点（图 7-25）。DNA 与核小体之间的这种联系是通过组蛋白和 DNA 间大量的氢键（约 40 个）所介导的。这些氢键大多位于蛋白质和 DNA

小沟附近磷酸二酯骨架上的氧原子之间，只有 7 个氢键位于蛋白质侧链和 DNA 小沟的碱基之间，所有这些氢键都在 DNA 的小沟里。

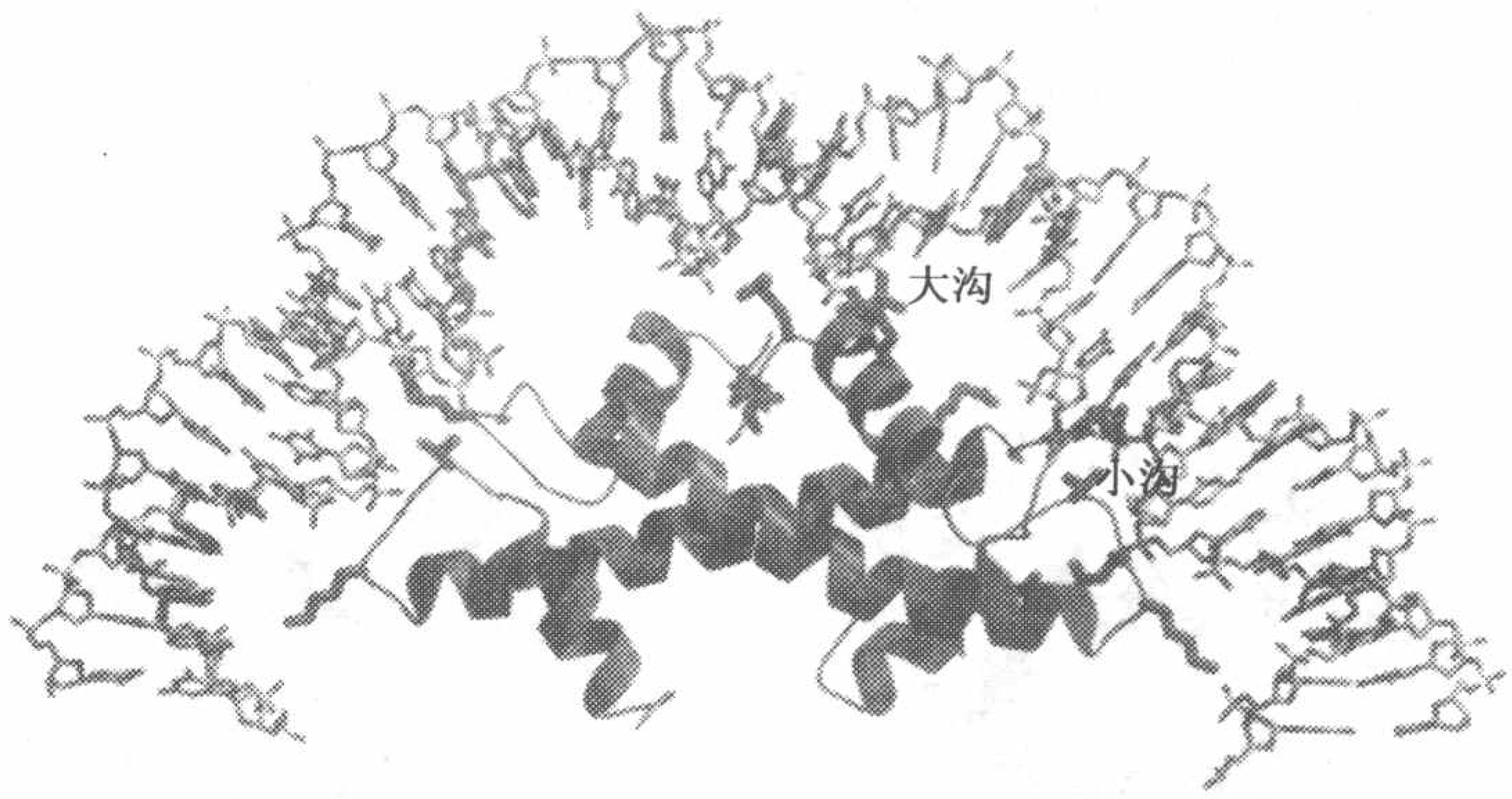


图 7-25　组蛋白与 DNA 间的接触位点。为了清楚起见，图中只显示了 DNA 与一个 H3 · H4 二聚体的相互作用。参与组蛋白与 DNA 相互作用的部分亚基用红色表示。注意这些区域聚集在 DNA 的小沟处。(Luger. 1997. Nature 389：251-260.) 镜像图用 BobScript、MolScript 和 Raster 3D 制作。

这些氢键中的大多数（一个典型的识别特异序列的 DNA 结合蛋白质仅有 20 个氢键与 DNA 作用）提供使 DNA 弯曲的驱动力。高度碱性的组蛋白通过屏蔽具抵抗 DNA 弯曲作用的磷酸基团的负电荷而将 DNA 弯曲。只有 DNA 弯曲使弯曲区内部的磷酸基团变为不适宜的邻近状态。组蛋白的高碱性特点还有助于两个相邻 DNA 螺旋紧密毗邻，这种毗邻对于 DNA 在组蛋白八聚体上多次缠绕是必要的。

组蛋白与 DNA 间的所有接触位点都与 DNA 的小沟或磷酸骨架相关。这个发现与组蛋白八聚体与 DNA 结合的非序列特异性的事实相一致。无论是 DNA 的小沟，还是磷酸骨架都不富含具碱基特异性的信息。而且，在与 DNA 小沟中的碱基所形成的 7 个氢键中，都不含有可区分 G：C 和 A：T 碱基对的元素（第 6 章，图 6-10）。

组蛋白的 N 端尾可稳定盘绕在八聚体上的 DNA

核小体的结构也能使我们对组蛋白 N 端尾巴的情况有所了解。4 个 H2B 和 H3 的尾巴从两个 DNA 螺旋间伸出。它们伸出的路径由两个相邻的小沟形成，在两个 DNA 螺旋之间形成了一个仅容一条多肽链的缝隙（图 7-26a）。值得注意的是，绕着组蛋白八聚体圆盘所伸出的 H2B 和 H3 尾巴之间的距离大致相等（H3 尾巴在大约 1 点和 11 点处伸出，H2B 尾巴在大约 4 点和 8 点处伸出）。H4 和 H2A 氨基端尾巴不是从两个 DNA 螺旋之间伸出，而是从 DNA 螺旋的“上面”或“下面”伸出来的（图 7-26a）。H4 尾巴位于核小体表面的 3 点和 9 点处，H2A 尾巴位于 5 点和 7 点处（图 7-26b）。从 DNA 螺旋之间或两侧伸出的组蛋白尾巴，其作用类似于一个螺丝上的凹槽，指导 DNA 以左手方式盘绕于组蛋白八聚体的圆盘上。正如我们在第 6 章中所讨论的一样，DNA 以左手方式缠绕使 DNA 形成负的超螺旋。组蛋白尾巴上最接近组蛋白圆盘的部分（不会被蛋白酶降解）当有 DNA 穿过时，也可在组蛋白和 DNA 之间形成许多氢键。

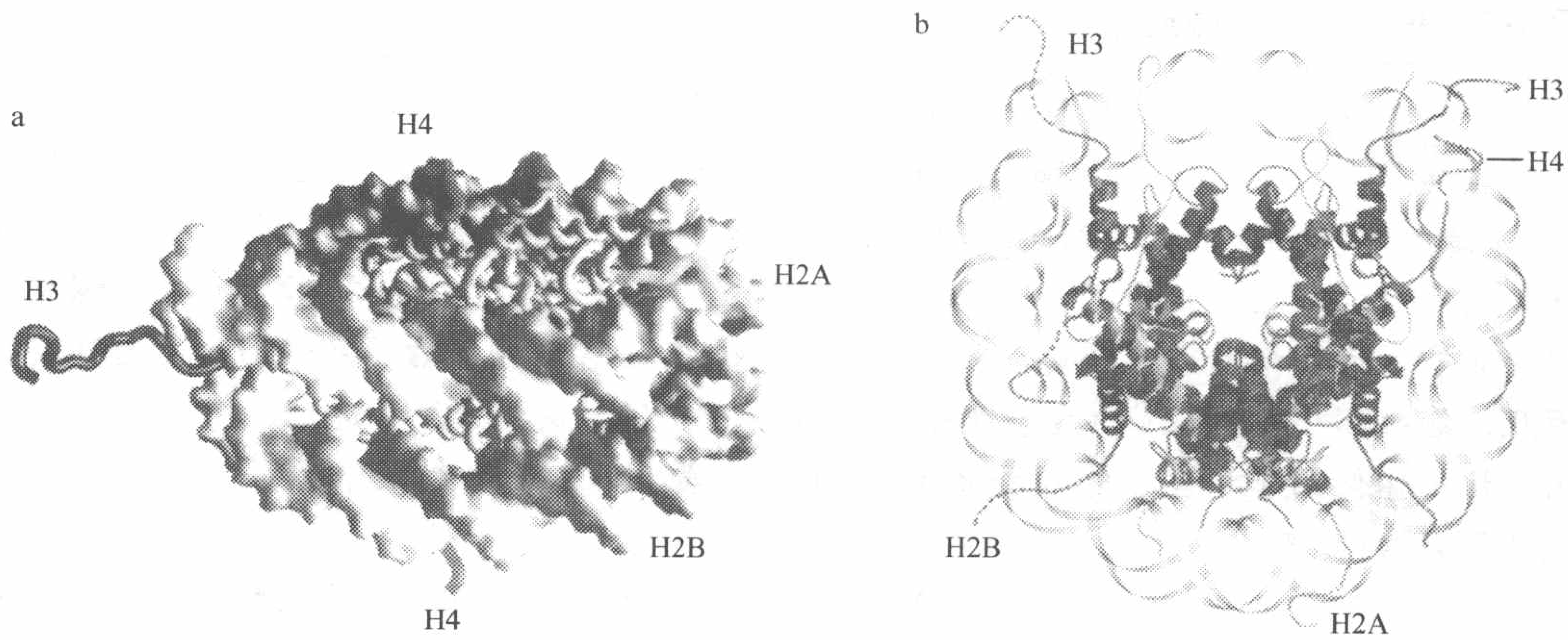


图 7-26 组蛋白尾巴从核小体核心的特定位置伸出。 a. 这个侧面图展示了从两个 DNA 螺旋间伸出的 H2B 和 H3 尾巴，相反，H4 和 H2A 尾巴从两个 DNA 螺旋的上面或下面伸出。(Luger K. et al. 1997. Nature 389:251-260.)示意图用 GRASP 绘制。b. 图中所示为组蛋白尾巴相对于 DNA 进出的位置，组蛋白尾巴可在 DNA 的很多相应位置上伸出。(Davey C. A. et al. 2002. J. Mol. Biol. 319:1097-1113.)镜像图用 BobScript、MolScript 和 Raster 3D 制作。

缠绕组蛋白核心的 DNA 呈负超螺旋特性

往共价闭环 DNA 模板上每加一个核小体，就能使相应 DNA 的连环数（linking number）改变约 1.2。由于 DNA 维持松弛状态主要依赖拓扑异构酶（topoisomerases），所以如果将核小体和 DNA 分离，被包装成核小体的 DNA 将形成负超螺旋。因此核小体可以被看成是维持或者稳定负超螺旋的结构。细胞为什么要保持这样一个负超螺旋的储库呢？已有很多例子说明该储库对细胞内 DNA 进行解链具有重要的促进作用，如 DNA 复制的超始，转录和重组。更重要的是，DNA 的负超螺旋有利于 DNA 的解旋（见图 6-17）。因此，核小体移除不仅增加了 DNA 的可接近性，还有利于附近 DNA 序列的解旋（框 7-2）。

如果核小体在真核生物细胞中具有储存负超螺旋的功能，那么是什么在原核生物中起相应作用呢？答案是多数原核生物的整个基因组都保持在负超螺旋状态。这种作用是由一种称作促旋酶（gyrase）的特殊的拓扑异构酶通过减少连环数来增加负超螺旋的方法来实验的。例如，在 E. coli 细胞中，促旋酶的作用结果是使整个基因组具有大约平均−0.07 的超螺旋密度。把负超螺旋导入松弛 DNA 是一种需能反应，因此，促旋酶只有在 ATP 存在的情况下才能把负超螺旋导入 DNA 中。在缺少 ATP 的情况下，促旋酶仅能使 DNA 松弛（例如，减少正超螺旋 DNA 的连环数）。

并不是所有的细菌都需要将它们的 DNA 保持在负超螺旋状态。适宜在极高温度条件下生长的细菌必须消耗额外的能量防止 DNA 由于热变性而解旋。这些生物有一种不同的拓扑异构酶，称作反促旋酶（reverse gyrase）。在 ATP 存在的情况下，反促旋酶可以增加松弛 DNA 的连环数。热变性通常会导致基因组的很多区域被解旋，反促旋酶通过维持基因组的正超螺旋结构逆转热变性的效果。

框 7-2 核小体和超螺旋密度

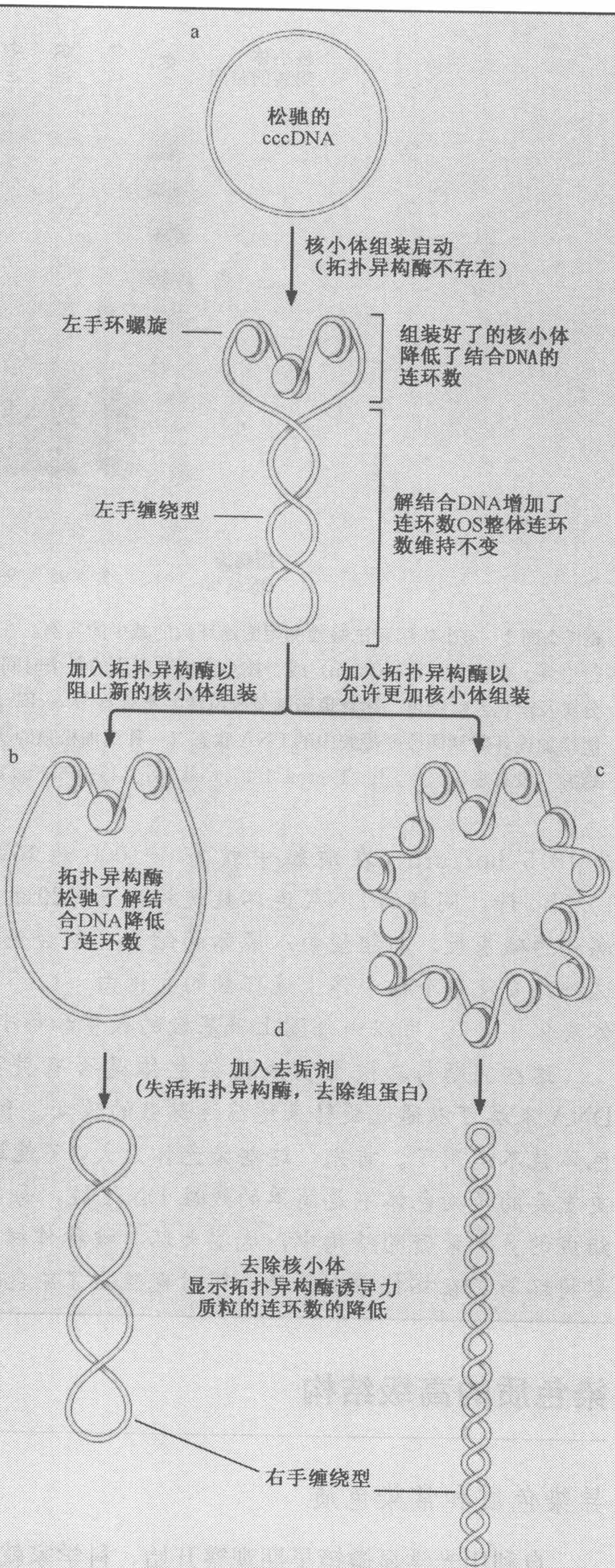
为什么核小体能改变自己含有 DNA 的拓扑结构呢？正如第 6 章所述，两种扭曲方式对 DNA 超螺旋的形成具有重要作用：环螺旋型的和缠绕型的（也称绞螺旋型的）。DNA 是以环螺旋方式缠绕到组蛋白八聚体上的。左手螺旋或右手螺旋决定了引入的超螺旋是正超螺旋还是负超螺旋（即相关 DNA 连环数的增加或减少等等）。对环螺旋型扭曲而言，左手螺旋引入负超螺旋（对缠绕型扭曲而言则相反右手螺旋引入负超螺旋）。因此，缠绕核小体 DNA 的左手螺旋环型扭曲减少了相关 DNA 的连环数。基于这个原因，核小体倾向于整合负超螺旋密度的 DNA，要用 DNA 组装核小体并且引入正超螺旋是很困难的。

用共价闭和环状 DNA（cccDNA）来组装很多核小体需要拓扑异构酶来适应与组蛋白结合的 DNA 的连环数的变化（见框 7-2 图 1）。如果缺少拓扑异构酶，cccDNA 组装成每一个核小体时结合时，需要非结合的 DNA（非核小体相关 DNA）来适应连环数的等量的增加（注意，在缺少拓扑异构酶的情况下，cccDNA 的总连环数是确定的）。因此，非结合 DNA 将会积累增加的连环数，增大正超螺旋的密度。非结合 DNA 的正超螺旋数越多，就越难组装到核小体里去。

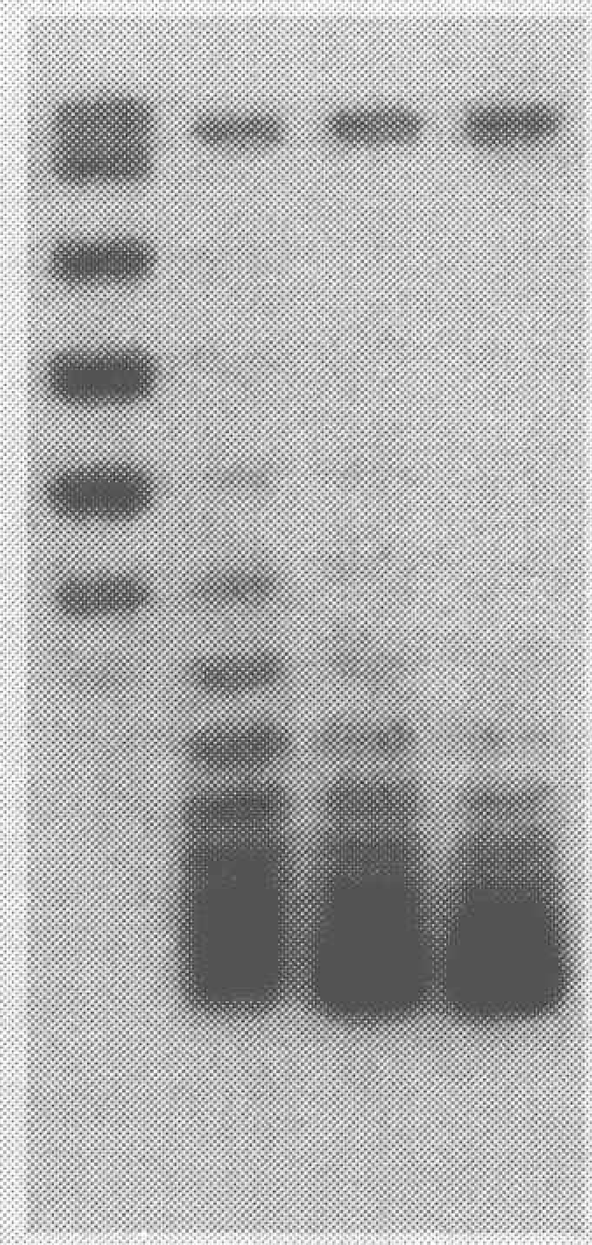
拓扑异构酶的加入能极大的促进核小体与 cccDNA 的结合。核小体组装时如果存在拓扑异构酶，它并不能作用于和核小体结合的 DNA，而是使非核小体结合 DNA 松驰，通过减少这些区域的连环数来降低正超螺旋的密度。拓扑异构酶通过维持非结合 DNA 的松驰状态，来促进组蛋白和 DNA 的结合以及更多的核小体的形成。值得一提的是，在质粒中的总体效应是随着核小体的不断被组装，连环数不断下降。

由拓扑异构酶引起的连环数减少，能够作为核小体组装过程的实验分析指标。该分析基于凝胶电泳能将松弛和超螺旋 cccDNA 区分开来（见图 6-27）的原理。第一步是在拓扑异构酶存在的条件下以 cccDNA 组装核小体。在适当的时间点往组装反应中加入强去污剂（如 SDS），使拓扑异构酶失活并将组蛋白从 DNA 上去除。得到的 DNA 通过凝胶电泳分离并鉴定其超螺旋特性。因为在组蛋白从 DNA 上去除的同时，拓扑异构酶也失活，保留了组装成核小体的 DNA 的连环数。一般来说，拓扑异构酶能把组装到每个核小体上的 cccDNA 减少 1.2 个连环数。因此，组装到核小体上的 cccDNA 越多，形成的负超螺旋数也越多（框 7-2 图 1c，d）。超螺旋的电泳迁移率较快，这在凝胶电泳中很容易被观察到（框 7-2 图 2）。

由于核小体 DNA 在组蛋白上缠绕 1.65 圈，环状共价闭合的质粒形成单个核小体的时候就会导致－1.65 圈的扭曲，所以会造成等量的连环数变化。就像上面所说的，当测量每个核小体对应的连环数时，得到的结果为－1.2 小于这个数。这种偏差称为“核小体连环数悖论”，而解决这一难题的关键是揭示核小体高溶解作用的晶体结构是在什么时候被溶解的。对组蛋白核心相关的 DNA 的深入分析表明：相对于裸 DNA 而言，每圈螺旋的碱基数减少了（从 10.5～10.2bp/turn）。每个螺旋的碱基数减少导致那段 DNA 的连环数增加。考虑到第 6 章中 cccDNA（10 500bp）这个例子。常规的 B 型 DNA



框 7-2 图 1　共价闭环 DNA（cccDNA）组装核小体需要拓扑异构酶。（a）在缺少拓扑异构酶的情况下，非结合 DNA 上的正超螺旋的积累抑制了核小体和 cccDNA 的组装。（b）拓扑异构酶存在的情况下，核小体装配并没有增加。示拓扑异构酶是如何通过减少连环数使非结合 DNA 松弛的。（c）拓扑异构酶存在的情况下，增加核小体装配。（d）同时去除组蛋白和灭活拓扑异构酶（例如，通过加入强去污剂）表明连环数的减少与核小体 DNA 相关。



框 7-2 图 2 核小体组装试验检测相应连环数的减少的实例。在拓扑异构酶存在的情况下，核小体以 cccDNA 进行组装。在组装开始（0 min）或者核小体组装反应的各个时间点加入去污剂，并且通过非变性琼脂糖凝胶电泳分离及溴化乙啶染色。尽管琼脂糖凝胶电泳并不能区分 cccDNA 是正超螺旋还是负超螺旋，但是 DNA 插入剂能增加连环数从而改变凝胶中的 DNA 状态（一种更加松弛的状态），根据这一点能够用来显示 cccDNA 的负超螺旋（没有显示）。经 Ito T. et al. 1997. Cell 90：145-155. Fig. 2c © Elsevier 许可使用.）

为 10.5 bp/turn，在质粒中就有＋1000 连环数（10 500/10.5）。相反，以每螺旋 10.2bp 计，同样的 DNA 连环数就约为＋1029 的（10 500/10.2）。因此，通过减少每圈螺旋的碱基数，与组蛋白八聚体的结合，就会使核小体结合 DNA 的连环数略微增加。这种改变使每个核小体上连环数的变化由－1.65 降到－1.2。这一差异（大约每个核小体变化＋0.4）可以由每圈上碱基数的数量和核小体 DNA 的长度的改变计算得到。

这些问题与真核生物线性染色体是否有关？短的线性片段与超螺旋无关，因为 DNA 末端可以通过旋转来适应连环数的改变。但对于非常大的真核生物细胞的线性染色体就不适用了。首先，这些染色体太大，不能通过快速旋转来消除 DNA 超螺旋变化。更重要的是染色体不是简单的线性 DNA 链，每条染色体 DNA 都被折叠到一个由大环组成的更加紧密的结构中，这些大环与被称作核骨架的蛋白结构拴在一起。这些附属装置将环与环在拓扑学上分开，同时也阻止了染色体 DNA 的自由旋转。

染色质的高级结构

异染色质和常染色质

自利用光学显微镜早期观察开始，科学家就发现染色体的结构并不是一致的。对染色体的早期研究将染色体区域分为两个部分：异染色质和常染色质。异染色质能被多种

染料染成深色、具有更为紧密的形态，而常染色质则具有相反的特征：染色浅而具有相对伸展的结构。随着分子生物学对基因和基因表达了解的深入，发现染色体的异染色质区域的基因表达量非常有限，而常染色质情区域的基因表达量较高。这一发现提示染色体的这两种不同染色质的结构与基因表达的总体水平有所关联。

尽管异染色质区域显示基因表达很少，但这并不意味这些区域不重要。正如我们在讨论基因表达时将学到的，关闭一个基因与启动一个基因可以说是同等重要的。再则，异染色质与包括端粒和着丝粒在内的特定染色体区域相关，对这两个染色体关键组分的功能非常重要。

经过若干年的研究，研究人员已对异染色质和常染色质结构有了分子水平上更加完整的理解。现在已清楚，这两种染色质中的DNA都被组装进核小体中。异染色质结构和常染色质结构的不同在于位于这两个不同染色体区域的核小体是如何被（或不被）组装进更大的组件之中。现在也很清楚，异染色质区由组装进高级结构的核小体DNA组成，并因此形成了基因表达的障碍。相反，常染色质区的核小体的组织程度要低得多。在下面的章节中，我们将讨论核小体是如何被组装进高级结构的。

组蛋白H1与核小体之间的连接DNA结合

一旦核小体形成，DNA包装的下一步就是和组蛋白H1的结合。与核心组蛋白相似，H1是一种小的、带正电荷的蛋白质（表7-5）。H1与核小体之间的连接DNA的相互作用，进一步加强了核小体与DNA间的紧密结合，增加了对核小体DNA的保护。这可以从对微球菌核酸酶消化的抵抗力增强所检测到。因此，除了核心组蛋白保护147bp的DNA之外，组蛋白H1与核小体的结合保护了额外的20bp DNA免于微球菌核酸酶的消化。

组蛋白H1具有不同寻常的特点，可与DNA双螺旋的两个不同的区域结合。通常，这两个区域是与核小体结合的同一DNA分子的一部分（图7-27）。相对于核小体而言，H1的结合位点是不对称的。H1的一个结合区是核小体一端的连接DNA。另一结合位点在与核小体结合的147bp DNA的中间（在二价轴上仅有的DNA双螺旋）。因此，上述的免于核酸酶消化的额外的DNA仅是核小体一侧的连接DNA。H1的结合使DNA的两个区域紧密靠近，增加了在组蛋白八聚体上紧密盘绕的DNA的长度。

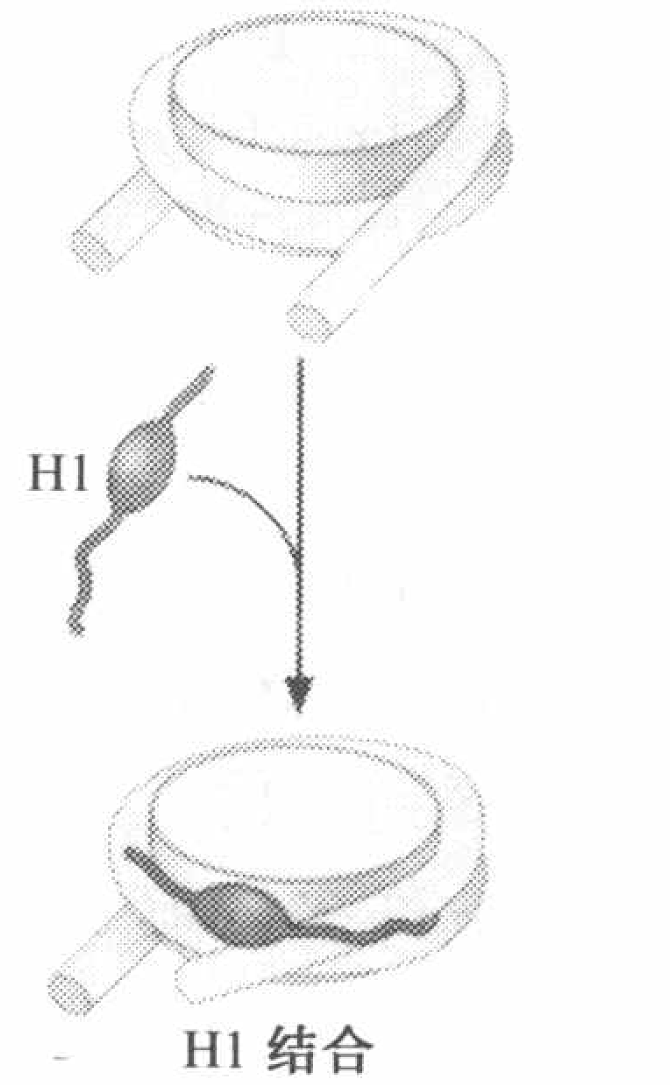


图7-27 组蛋白H1结合两个DNA螺旋。就组蛋白H1与核小体的相互作用而言，组蛋白H1分别与核小体一端的连接DNA及与核小体结合的DNA中部的DNA螺旋结合（即核心组蛋白八聚体结合的147bp DNA的中间）。

H1的结合使DNA进出核小体的角度更为固定（图7-28）。通过电子显微镜可以观察到，这种效果使核小体DNA呈现出独特的锯齿形外观。DNA进出的角度视条件（包括盐浓度、pH和其他蛋白质的存在）变化很大。如果我们假设这些角度相对于二价轴是约20°，这将导致核小体交替存在于组蛋白H1结合的连接DNA中心区的两侧（图7-29）。

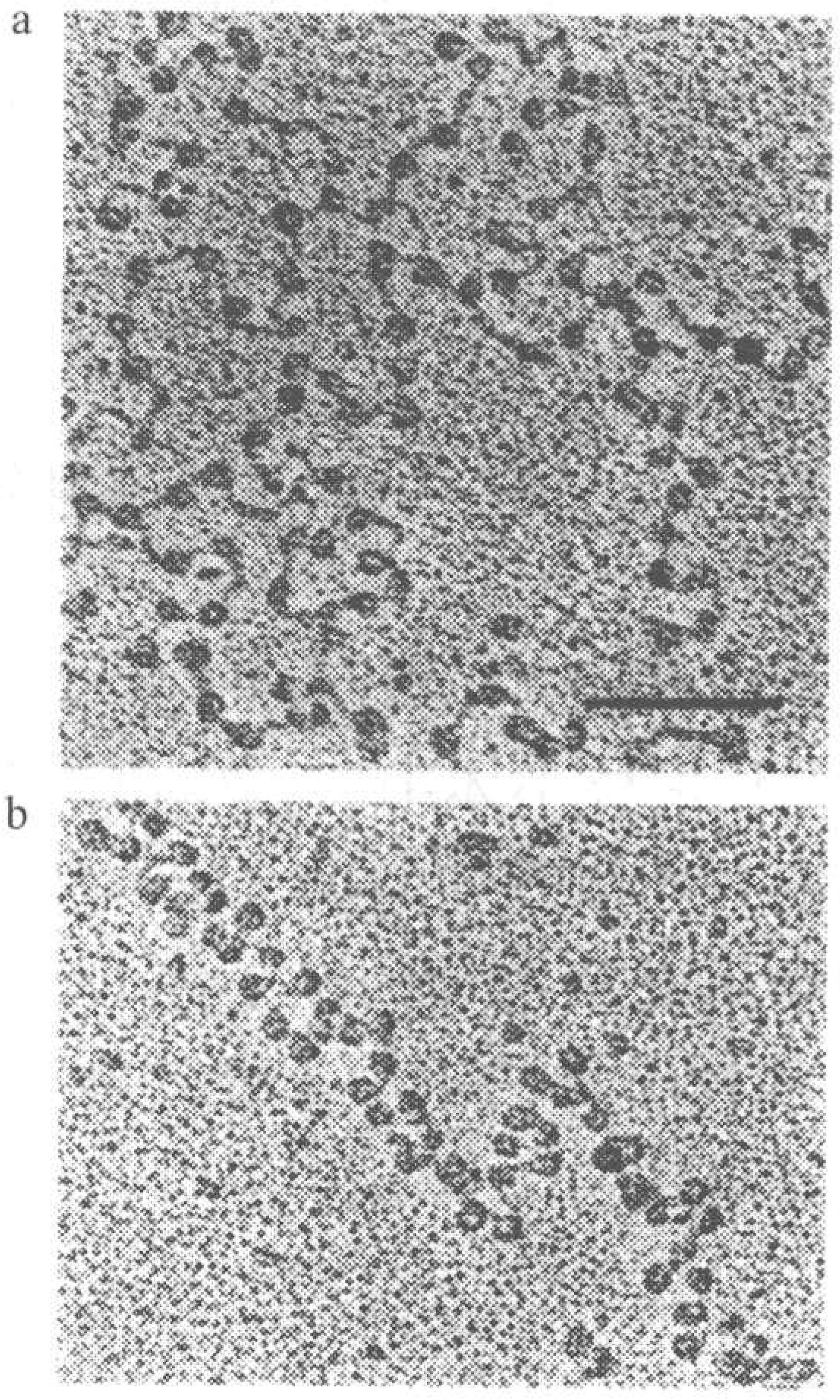


图 7-28　H1 的结合导致核小体 DNA 的进一步凝聚。这两幅照片是在电子显微镜下显示的组蛋白 H1 存在（a）或缺失（b）时核小体 DNA。注意在组蛋白 H1 存在时，核小体 DNA 的凝聚更紧密且结构更固定。（来源：Thoma et al. J. Cell Biology，83：410，Figs 4 & 6. © Rockefeller University Press.）

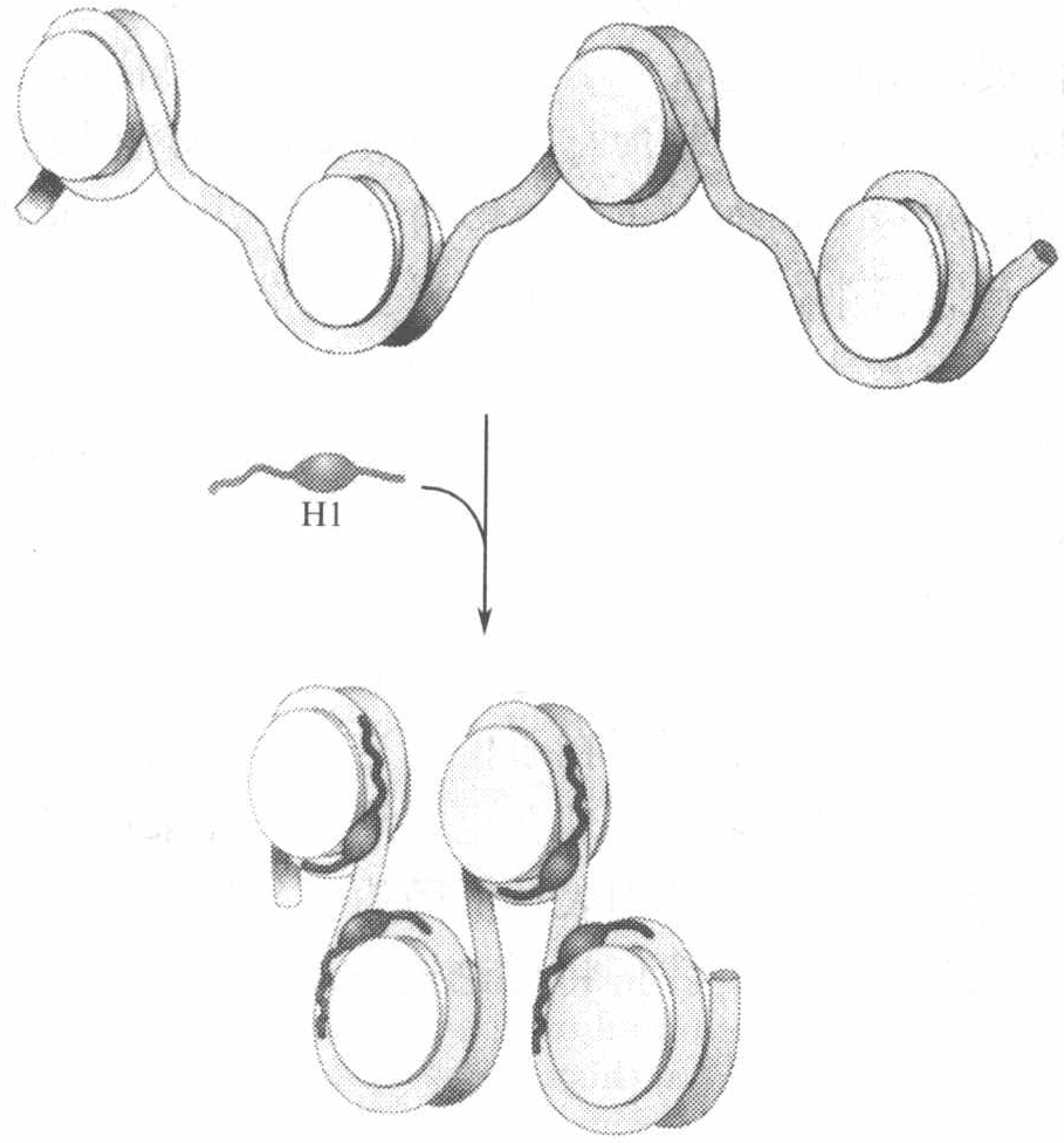


图 7-29　组蛋白 H1 使 DNA 更紧密的盘绕于核小体。这两幅图是组蛋白 H1 存在或缺失条件下 DNA 盘绕核小体情况的比较。组蛋白 H1 能与每个核小体作用，结合于连接 DNA 和核小体结合 DNA 中部的 DNA 螺旋。

核小体束能形成更复杂的结构：30nm 的纤丝

组蛋白 H1 的结合稳定了染色质的高级结构。在试管中，当盐浓度增加时，组蛋白 H1 的加入导致核小体 DNA 形成一条 30nm 的纤丝（30nm fiber）。这个结构也能在体内观察到，代表了 DNA 凝聚的下一层次。更重要的是，DNA 整合到这种纤丝内使其不易被许多依赖 DNA 的酶（如 RNA 聚合酶）接近。

关于 30nm 纤丝的结构，目前有两种模型。在螺线管模型（solenoid model）中，核小体 DNA 形成一个超螺旋，每圈大约包含 6 个核小体（图 7-18a）。电镜和 X 射线衍射研究都支持这个结构。研究表明，30nm 的纤丝有一个大约 11nm 的螺旋斜度。这也是核小体圆盘的大约直径，表明 30nm 纤丝是由核小体圆盘堆叠成螺旋状所构成（图 7-30a）。在这个模型中，组蛋白八聚体圆盘的平面间相互连接，而核小体的 DNA 表面形成超螺旋外部易接近的表面。连接 DNA 被埋藏在超螺旋的中心，但从不穿过纤丝轴。而且，当 DNA 从一个核小体移到另一核小体时，连接 DNA 环绕着中心轴。

30nm 纤丝的另一模型是锯齿模型（图 7-30b）。这个模型的基础是结合 H1 后的核小体呈锯齿状。在这种模型下，30nm 纤丝是这些锯齿形核小体束的压缩结构。新近的

以4个核小体的DNA单一分子的X衍射和对弹簧状的单一30nm纤丝的分析支持这个模型。与螺线管模型不同，锯齿状的构象要求连接DNA以相对笔直的方式穿过纤丝的中心轴（图7-30b）。因此，较长的连接DNA有利于这种构象的形成。因为连接DNA的平均长度在不同物种间有变化（表7-4），所以30nm纤丝的结构并不总是相同，这两个模型也许都是对的。

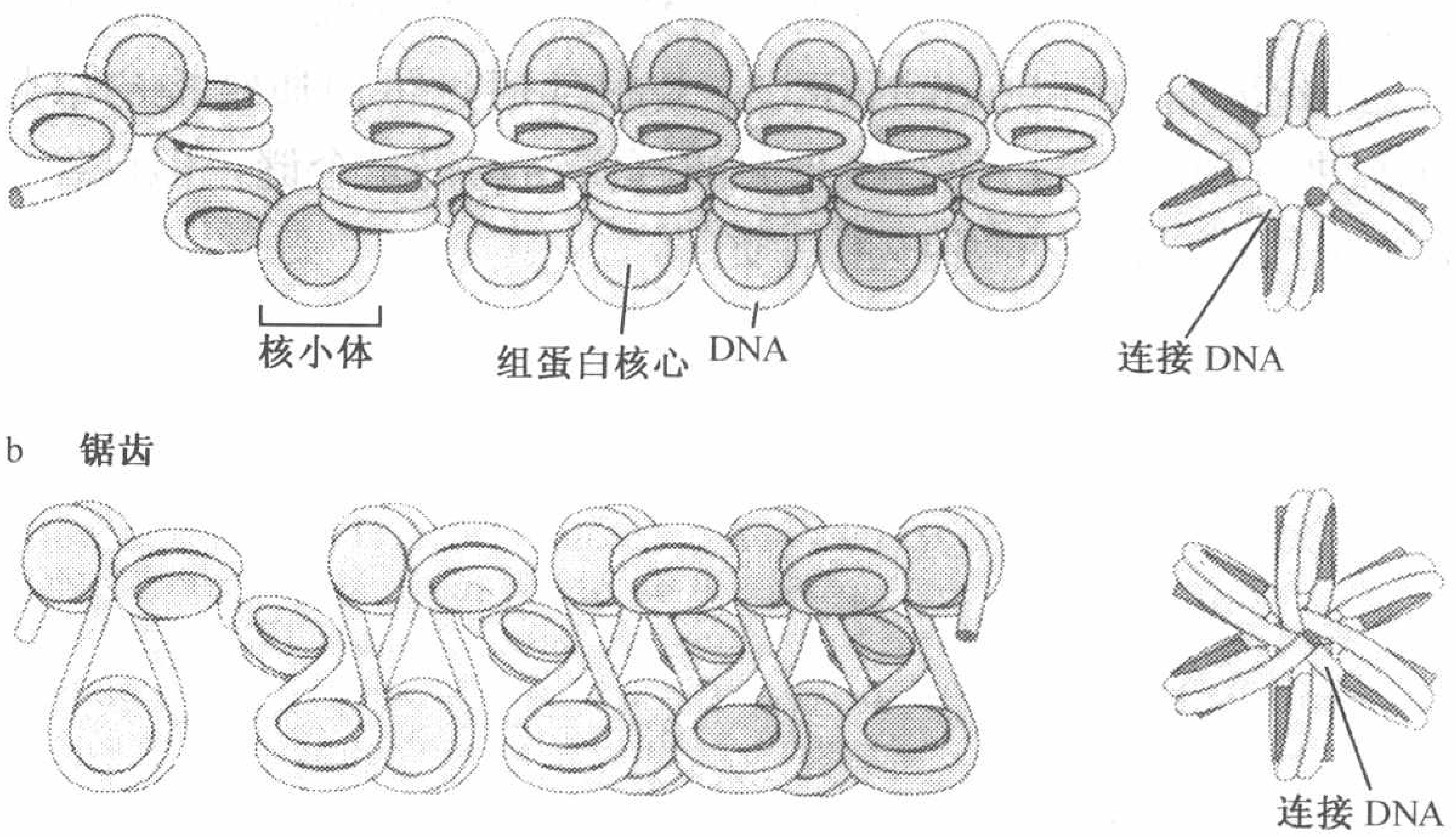


图7-30 30nm染色质纤丝的两种模型。在a、b两组中，左边为纤丝的侧面，右边为沿纤丝的中心轴的下观图。a. 螺线管模型。注意连接DNA不穿过超螺旋的中心轴且进出核小体的位点和侧面相对不易靠近。b. 锯齿模型。在这种模型中，连接DNA经常穿过纤丝的中心轴，侧面甚至进出位点更易靠近。（来源：Pollard T. and Earnshaw W. 2002. Cell Biology，1st edition，p. 202，Fig. 13-6. © Elsevier）

组蛋白N端尾巴对于30nm纤丝的形成是必需的

缺少N端尾巴的核心组蛋白不能形成30nm纤丝。这个尾巴最可能的作用是通过它与相邻核小体的相互作用稳定30nm纤丝。核小体的三维晶体结构支持这种模型。核小体三维结构显示，在晶格中，H2A、H3和H4的N端尾巴各自与相邻的核小体相互作用（图7-31）。最近的研究表明，组蛋白H4带正电荷的N端与H2A的组蛋白折叠压的负电荷压，对于30nm纤丝的形成特别重要。与H4尾巴相互作用的H2A残基在许多真核生物中是保守的，但这些残基不参与DNA的结合或组蛋白八聚体的形成，一种可能性是H2A的这些区域是保守地介导核小体相互之间与H4尾巴的相互

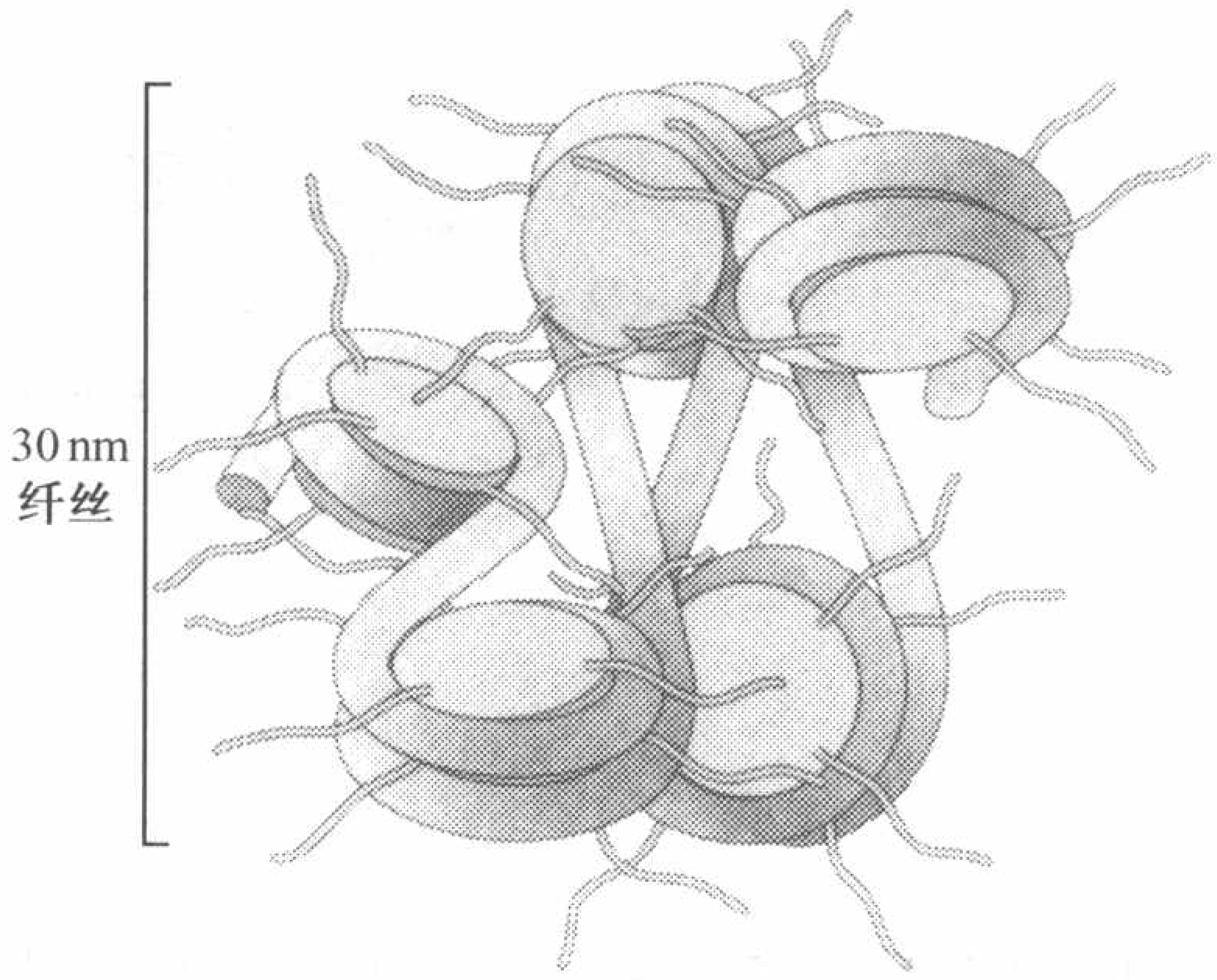


图7-31 组蛋白N端尾巴稳定30nm纤丝的推测模型。在这个模型中，30nm纤丝用锯齿模型表示。几种不同的核心组蛋白与尾巴的相互作用是可能的。这里描绘了每个交替的组蛋白间的相互作用，但这种相互作用也能在相邻的或更远的组蛋白间发生。

作用。我们将在下面看到，在细胞中组蛋白尾巴经常是修饰的靶标。这些修饰可能影响 30nm 纤丝和其他核小体高级结构的形成能力。

DNA 的进一步凝聚与核小体 DNA 的大环有关

DNA 包装为核小体和 30nm 纤丝共同导致 DNA 的线性长度压缩了大约 40 倍。但这种 1～2m 长的 DNA 仍然不能适合大约 10^{-5}m 长的核。为了使 DNA 更为压缩，必须对 30nm 纤丝进行进一步的折叠。尽管尚不清楚这种折叠的确切结构，但是一个流行的模型推测 30nm 纤丝形成 40～90kb 的环，在其根部通过蛋白质结构的核骨架（nuclear scaffold）连接起来（图 7-32）。虽然核骨架的真实结构仍是个谜，但已经研究出一些方法来鉴定作为这个结构一部分的那些蛋白质。

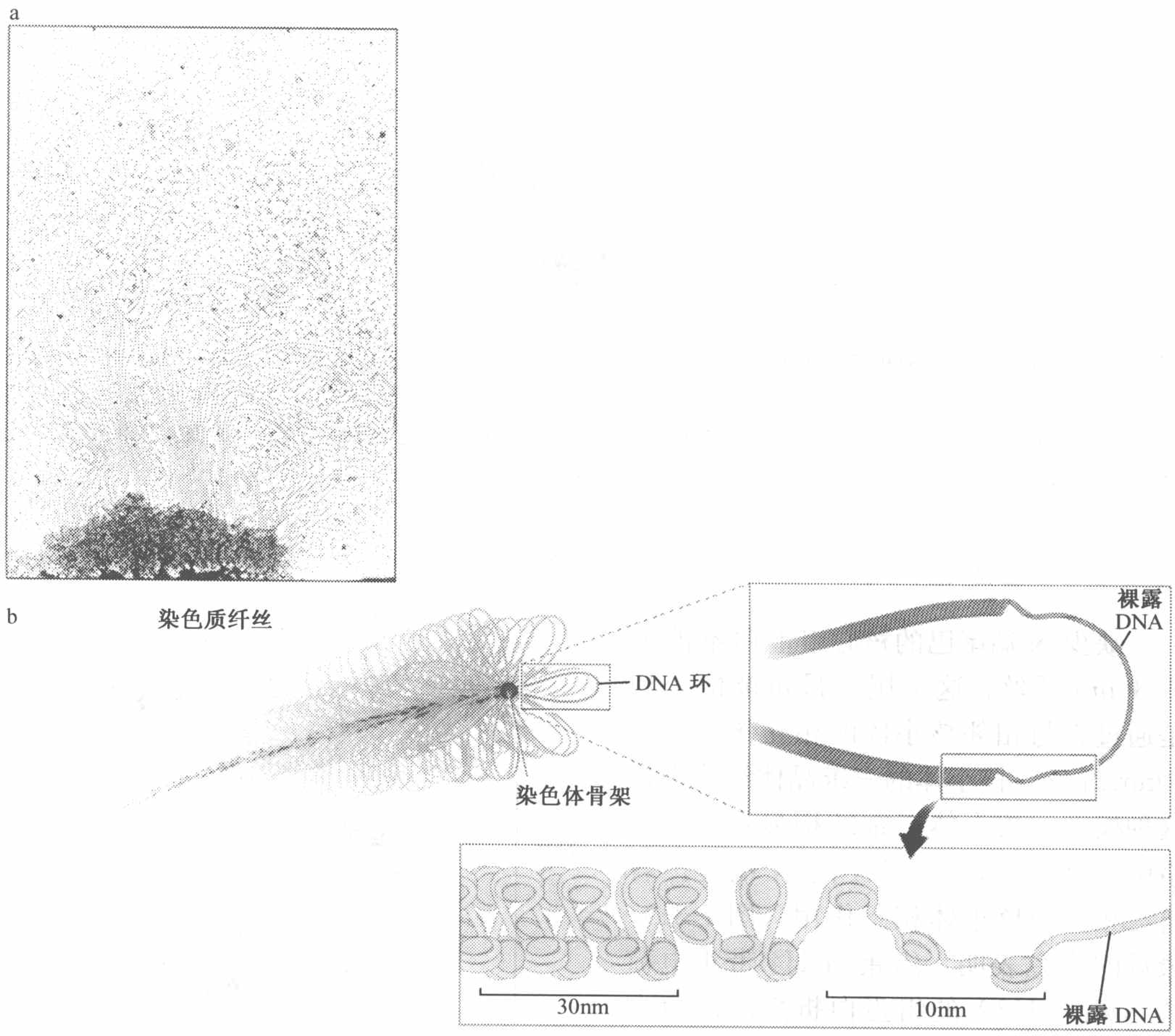


图 7-32 染色质的高级结构。 a. 一个透射电子显微照片显示染色质从染色体的中心结构处伸出。电子密集的区域是核骨架，在真核细胞中它负责将大量的染色体 DNA 组织起来。棒状物代表 200nm。b. 一个真核染色体的结构模型，显示大多数 DNA 包装进 30nm 纤丝的大环且其根部结合于核骨架上。而有活性的 DNA 位点（如转录或 DNA 复制位点）以 10nm 纤丝甚至裸露 DNA 的形式存在。（来源：a，Courtesy of J. R. Paulson and U. K. Laemmli.）

目前已发现有两类蛋白参与核骨架的形成。一类是拓扑异构酶Ⅱ（Topo Ⅱ），大量存在于核骨架的制备物和纯化的有丝分裂染色体中。用能够导致 DNA 在 Topo Ⅱ的 DNA 结合位点发生断裂的药物处理细胞，得到长约 50kb 的 DNA 片段。这与用有限的核酸酶消化染色体所得到的片段长度相似，表明Topo Ⅱ可能是参与形成这种根部相连的环状 DNA 的机制的一部分。此外，每个环的根部的 ToPⅡ可以保证这些环在拓扑学上的相互分离。

SMC 蛋白也是核骨架中含量丰富的组成成分。正如我们先前所讨论的（染色体复制和分离部分），SMC 蛋白是染色体复制后进行凝聚和维持姐妹染色体联会机制的关键组成成分。这些蛋白质通过与核骨架的结合，为其与染色体 DNA 的相互作用提供了内在的基础，从而增强它们的功能。

组蛋白的变构体影响核小体功能

核心组蛋白是最保守的真核蛋白质之一。因此，由这些蛋白质形成的核小体在全部真核生物中是非常相似的。但是在真核细胞中发现了很多组蛋白变构体。这些组蛋白能替代 4 种标准的组蛋白的一种，形成了替代的核小体。这类核小体可能用来划分染色体的特定区域，或者赋予核小体特定的功能。例如，H2A. X 是 H2A 的一种变构体，广泛分布于真核核小体中，当染色体 DNA 发生断裂的时候（称为双链断裂），邻近断点的 H_2A. X 会被磷酸化。这种被磷酸化后的 H2A. X 可以被 DNA 修复酶特异地识别，并将修复酶定位到 DNA 损伤位点。

另一个组蛋白变构体 CENP-A，与含有着丝粒 DNA 的核小体结合。在这个染色体区域，CENP-A 替代了核小体中的组蛋白 H3 亚基。这些核小体整合进动粒，介导染色体与有丝分裂纺锤体的黏附（图 7-12）。与 H3 相比，CENP-A 含有较长的 N 端尾巴。但含有另外的相似的组蛋白折叠结构域，CENP-A 的加入不大可能改变核小体的核心结构。然而，CENP-A 延长的尾巴可能给动粒的其他蛋白质组分提供新的结合位点（图 7-33）。按照这种假说，CENP-A 的丢失可能干扰动粒组份与着丝粒 DNA 的相互作用。

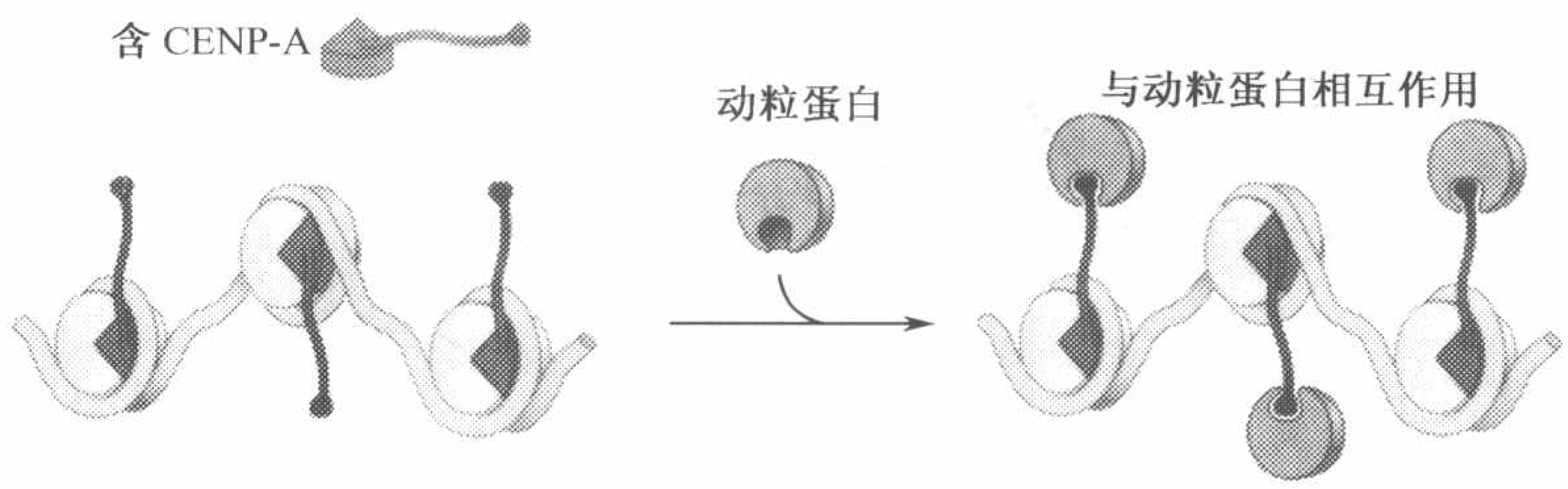


图 7-33　组蛋白变构体掺入引起的染色质改变。CENP-A 替代组蛋白 H3 可能为一个或多个动粒的组分提供了结合位点。

染色质结构的调控

DNA 与组蛋白八聚体的相互作用是一动态的过程

正如我们将在第 17 章所要详细讲述的，DNA 包装进核小体对基因组的表达有重要的影响。在许多情况下，能够靠近特定的 DNA 区域所需的关键步骤是将核小体去除或者使其与 DNA 的结合不那么紧密。为适应这个需要，组蛋白八聚体与 DNA 的结合是动态的。此外，有一些因子能作用于核小体，增加或降低这种动态结合。这些特征共同保证了核小体的位置和 DNA 相互作用可以发生改变，以适应对 DNA 易接近性经常变化的需要。

与所有非共价键介导的相互作用相似，任何特定的 DNA 区段与组蛋白八聚体的结合都不是永久的：任何 DNA 区段都将不时从与八聚体的紧密相互作用中短暂地释放出来。这种释放类似于 DNA 双螺旋的暂时解链（与我们在第 6 章中讨论的一致）。由于许多 DNA 结合蛋白更易结合没有组蛋白的 DNA，因此 DNA 与组蛋白核心结构结合的动态特点十分重要。只有从组蛋白八聚体上释放出来，或者结合到连接 DNA 上，即 DNA 上没有核小体的区段，这些蛋白质才能识别 DNA 上的结合位点。

DNA 从核小体上间歇性地、自发性地释放的结果，是蛋白质与其 DNA 结合位点接近的概率为 1/50 至 1/100 000，概率的大小取决于结合位点在核小体中的位置，越处于中心的结合位点越不易被接近。因此，在与核小体紧密结合的 147bp DNA 上的 73 位点附近的结合位点是最不易接近的位点，而在核小体 DNA 两端（1 或 147 位点）的结合位点最易于接近。这些发现表明 DNA 暴露的机制归因于 DNA 从核小体上解绕，而不是简单地从组蛋白八聚体表面伸出（图 7-34）。值得注意的是，这些研究都是在试管里用核小体的群体来进行的，与细胞中 DNA 从核小体上解绕的能力可能有所不同，因为在细胞中，长长

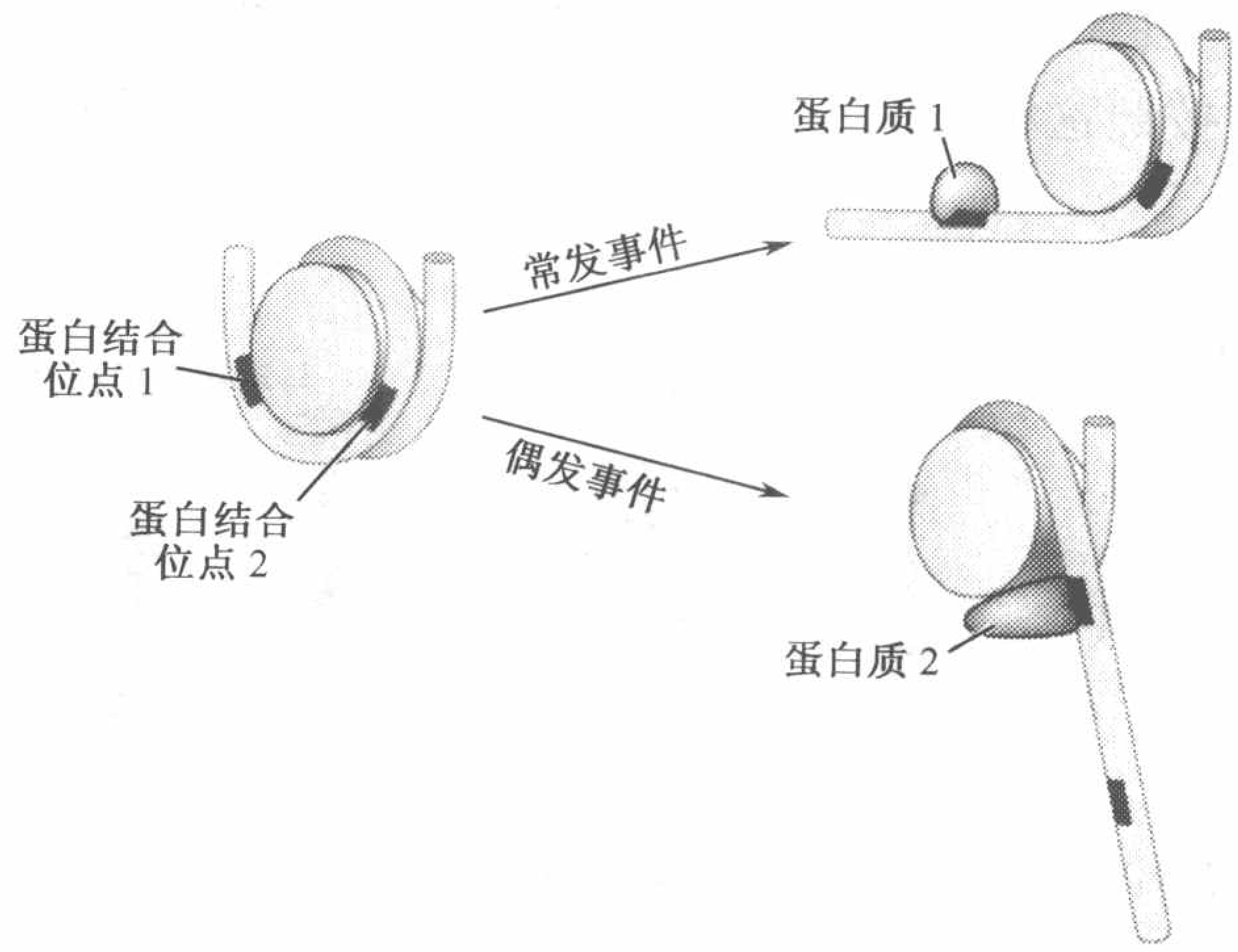


图 7-34 接近与核小体结合的 DNA 的模型。对具序列特异性的 DNA 结合蛋白与核小体结合能力的研究表明，DNA 从核小体上的解绕对于 DNA 的易接近性起决定作用。因此，最靠近核小体进出位点处的 DNA 结合位点是最易接近的，而最靠近结合 DNA 中部的位点是最不易少被接近的。

的 DNA 要组装到很多相邻的核小体（被称为核小体集束）里去。与 H1 的结合以及核小体组装成 30nm 的纤丝，都将改变这些概率。然而，核小体的动态结构表明，与 X-射线结晶学揭示相似的核小体结构只是短暂性的，而不是大部分时间里的其他构型。

核小体重塑复合体有助于核小体的运动

除了核小体所呈现的内在的动态性之外，组蛋白和 DNA 相互作用的稳定性还受到蛋白质大复合体，称为核小体重塑复合体的影响。这些多蛋白复合体有助于以 ATP 水解为能量的核小体-DNA 相互作用或核小体位置的改变。被这些酶催化的核小体改变，有两种基本的类型（图 7-35）。所有核小体重塑复合体都能催化 DNA 沿着组蛋白八聚体表面的“滑动（sliding）”。另一种类型的核小体重塑复合体能够催化第二种更为极端的变化，使组蛋白八聚体能从一个 DNA 螺旋向另一个螺旋“转移（transfer）”。

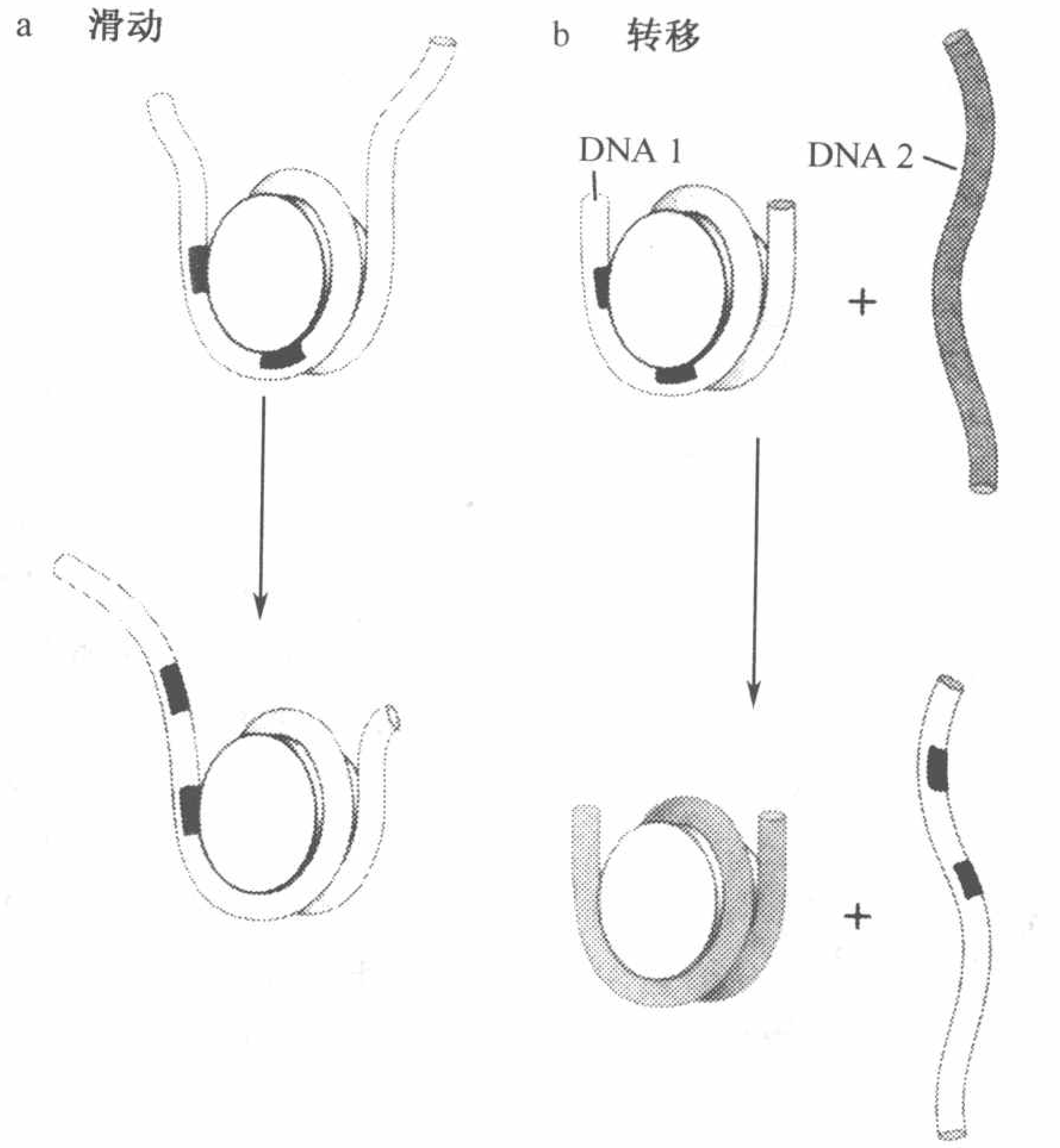


图 7-35 核小体重塑催化核小体的移动。 a. 核小体沿 DNA 分子滑动，暴露出 DNA 结合蛋白的结合位点。b. 核小体也可以从一条 DNA 链转移到另一条链上。

最近的研究已经揭示核小体重塑复合体是如何将 DNA 在组蛋白八聚体上移动的。每一个这样的多亚基酶都有一个 ATP 水解亚基，能够在双链 DNA 上定向移动（也称为移位）。就核小体重塑复合体来说，这些酶被认为能紧密结合到组蛋白八聚体上，并能将 DNA 相对于组蛋白移位。由于同时将既有的 DNA 都相对于组蛋白移位在能量上是困难的，有人提出 ATP 水解亚基只能在一次移动一部分 DNA，结果是导致 DNA 在组蛋白八聚体表面上的“尺蠖样（inchworm-like）”运动（图 7-36）。重要的是要记住，DNA 序列与组蛋白八聚体的相互作用的亲和性是基本等同的。这样，可以把 DNA 分子沿着组蛋白八聚体的滑动看成是不同的能量等同状态的与组蛋白八聚体结合，核小体重塑复合物使 DNA 更为容易地接近这些不同的状态。

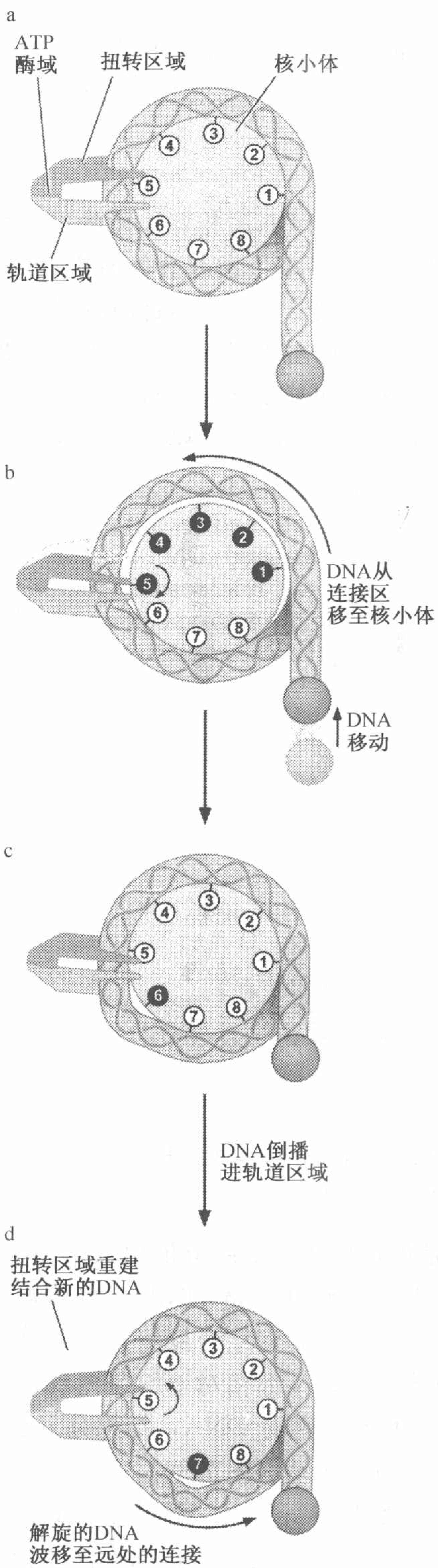


图 7-36 核小体重塑复合体俯化的核小体DNA滑动的一个模型 a. 据这一模型，核小体重塑复合体的ATP水解亚基结合到距中心二价轴两个螺固的核小体DNA上（如与核小体的147 bp中的52位）。核小体重塑复合物的其他亚基都与组蛋白紧密结合。图示DNA和组蛋白从二价轴到最邻近的未结合DNA的每一个接触点（每一螺周一个接触点，14个中的7个）。b. ATP依赖的DNA移位活性，使核小体重塑复合体首先将DNA从最近邻的连接区拉入核小体。ATP水解亚基和连接DNA之间的5个接触点被打断（被打断的接触点为黑色，完整的接触点为白色）。c. 被打断的接触点以移位的DNA重新形成，使紧靠ATP水解亚基（5和6接触点之间）与太多的DNA结合。d. 为解脱这一约束，这个模型提出其他序列像“波（wave）”一样沿着组蛋白的表面移动，每一次打断一个接触点（先是6，然后是7接触点），直到所有接触点都以它们之间适量的DNA重新形成。此时，与组蛋白结合的DNA不再过量，核小体已经在DNA上就位。（经许可修改自 Saha A. et al. 2006. Nat. Rev. Mol. Cell Biol. 7：437-447，Fig. 4a. © Macmillan.）

在任一细胞中都有多种类型的核小体重塑复合体（表 7-6）。它们至少有 2 个亚基，最多有 10 个亚基。正如前文所描述的（见图 7-36），这些复合体中的每一种都含有一个推进 DNA 移动的 ATP 水解亚基。尽管在不同的复合体中 ATP 水解亚基是相似的，但是与每一个复合物结合的其他不同亚基可以调节其功能。例如，这些复合体包含能将其定位于特定染色体位置上的亚基。在一些情况下，这种定位由重塑复合体亚基和 DNA 结合的转录因子间的相互作用所介导（图 7-37）。在其他情况下，结合到经特异修饰的组蛋白 N 端尾巴也可以介导定位（通过铬功能域（chvomodomains）或溴功能（bromodomains），将在下面讲到）。

表 7-6　ATP 依赖性核小体重塑复合体

类型	亚基数目	组蛋白结合域	滑动	转移
SWI/SNF	8—11	溴功能域	有	有
ISWI	2—4	溴功能域，SANT 域，PHD 指纹	有	无
CHD	8—10	铬功能域，PHD 指纹	有	无
SWR1	12—14	溴功能域，SANT 域	有	未定
INO80	10—12	无	有	未定

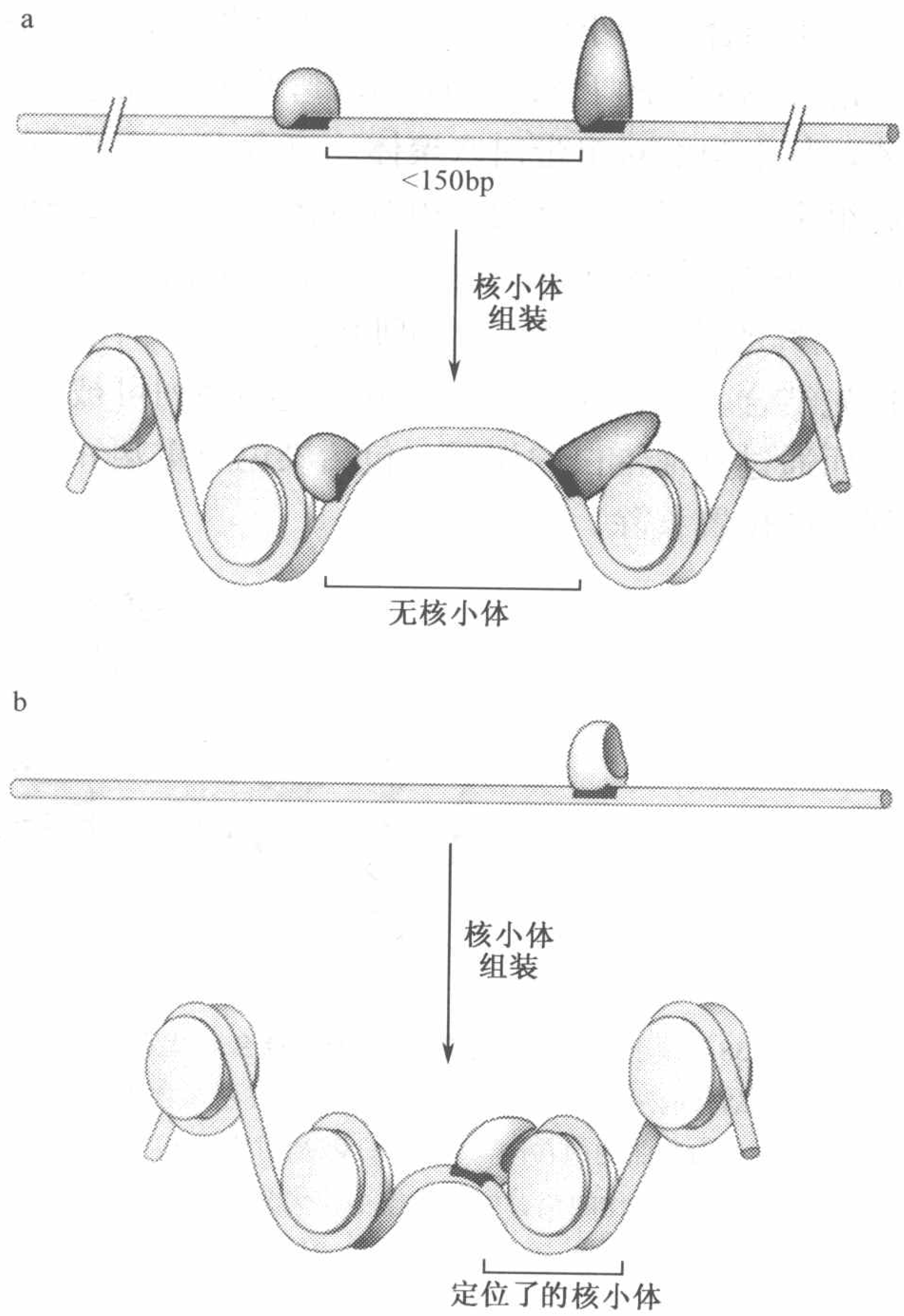


图 7-37　两个依赖于 DNA 结合蛋白的核小体定位的模型。a. 许多 DNA 结合蛋白与 DNA 的结合和同一 DNA 与组蛋白八聚体的结合是不兼容的。因为核小体的形成需要多于 147bp 的 DNA，如果两个与 DNA 结合的结合因子间的距离少于 147bp，那么二者之间的 DNA 不能组装进核小体。b. 一些 DNA 结合蛋白具有与核小体结合的能力。这些蛋白质一旦与 DNA 结合，将立即有助于核小体与相邻的蛋白质的 DNA 结合位点的结合。

某些核小体处于特定的位置：核小体的定位

由于核小体与 DNA 的相互作用是动态的，并且没有序列特异性，大多数核小体的位置是不固定的。但在有些情况，对核小体的位置，或称为核小体的定位（positioning），进行限制是有利的。通常，核小体的定位允许调节蛋白的 DNA 结合位点处于易接近的连接 DNA 区。在多数情况下，这些没有核小体结合的 DNA 区段要足够大，以使多个调节蛋白的结合位置易于被接近。

核小体定位可由 DNA 结合蛋白或特定的 DNA 序列指导。在细胞中，最经常的方式是核小体与 DNA 结合蛋白间的竞争。正如许多蛋白质不能与核小体内的 DNA 结合一样，当蛋白质与 DNA 结合后，就阻止了核心组蛋白与这段 DNA 随后的结合。如果两个 DNA 结合蛋白与 DNA 结合的位点间的距离小于 DNA 组装成一个核小体所需的最小长度（约 150bp），那么这两个蛋白质之间的 DNA 将不与核小体结合（图 7-37a）。相邻的 DNA 与额外的蛋白质的结合能进一步增加不与核小体结合的区域的长度。除了这种依赖于蛋白质的核小体定位的抑制机制外，一些 DNA 结合蛋白与相邻的核小体紧密作用，导致核小体立即优先地在与这些蛋白质相邻的位置组装（图 7-37b）。

核小体定位的另一种方式涉及特定的 DNA 序列，这些序列对核小体有高度的亲和性。由于结合在核小体上的 DNA 是弯曲的，所以核小体优先在 DNA 易弯曲的区域形成。富含 A∶T 的 DNA 有向小沟弯曲的内在倾向。因此，富含 A∶T 的 DNA 有利于组装，其小沟面对组蛋白八聚体。富含 G∶C 的 DNA 则有相反的倾向，因此，当小沟背离组蛋白八聚体时有利于组装（图 7-38）。每个核小体都试图使富含 A∶T 和富含 G∶C 序列的这种排列方式最大化，最近对酵母（*S. cerevisiae*）核小体定位的研究揭示，高达 50%的核小体严格定位可归因于这些核小体的组蛋白核心对其所包含的 DNA 序列的优先结合。值得注意的是，虽然这些序列可被优先结合，但却不是核小体组装所必需的，而通过其他蛋白的作用，包括染色质重塑和转录因子结合，还可将核小体从这些它们偏好的位置移除。

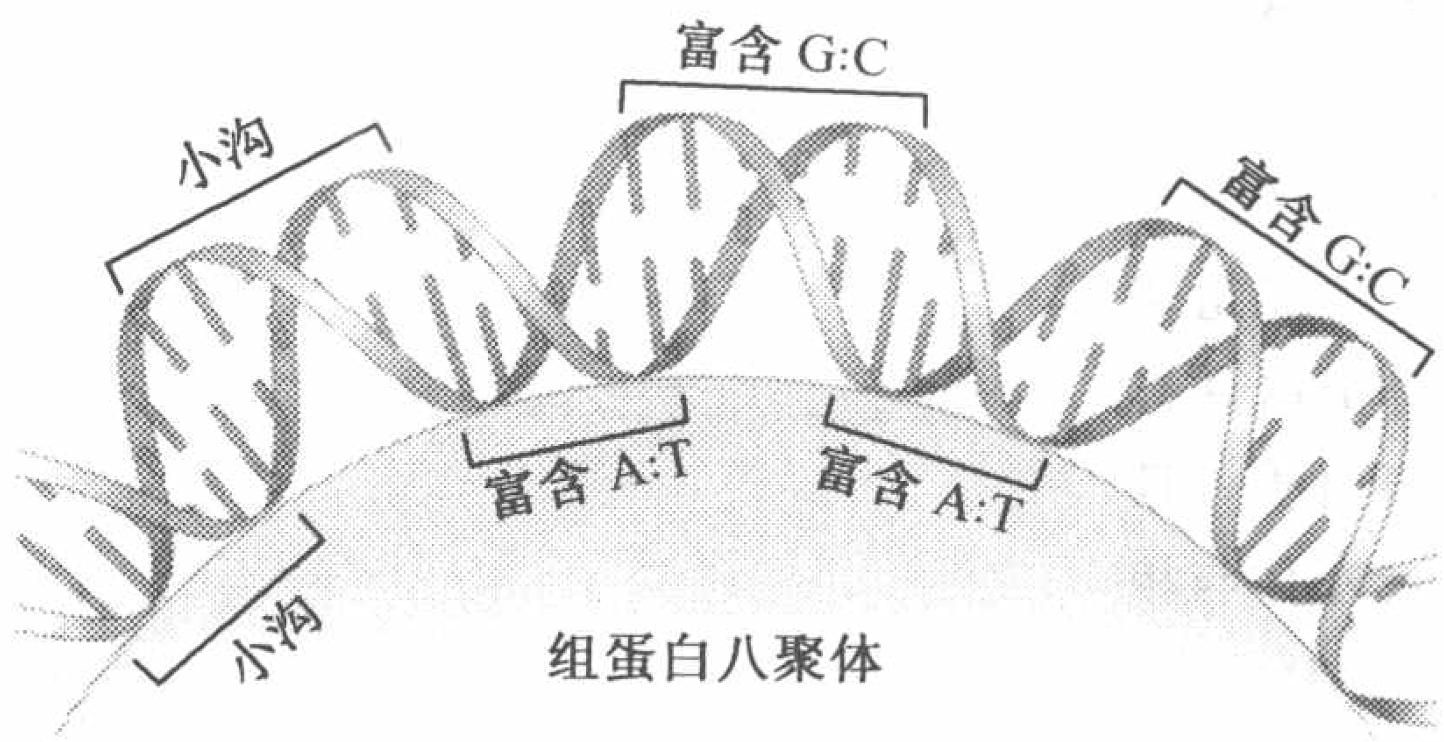


图 7-38　核小体倾向于与弯曲的 DNA 结合。特定的 DNA 序列能定位核小体。由于 DNA 在与核小体结合的过程中会剧烈弯曲，所以具有内在弯曲倾向的 DNA 序列可定位核小体。A∶T碱基对有向小沟弯曲的内在倾向，G∶C 碱基对则有相反的倾向。富含 A∶T 和 G∶C 的序列之间以约 5bp 为周期变化的交替序列是核小体优先的结合位点。（来源：Alberts B. et al. 2002. Molecular biology of the cell，4th edition，p. 211，f4-28. © Garland Science/Taylor & Francis LLC.）

核小体定位的这些机制会影响基因组中核小体的组织。尽管如此，大多数核小体并没有严格的定位。严格定位的核小体经常在指导转录起始的位点发现，这将在真核转录的章节（第 12 章和第 17 章）中讨论，虽然我们已经讨论了定位主要是作为保证调控 DNA 序列的可接近性的一种手段，一个定位的核小体仅需通过定位在与某一特定位点有重叠的位置上，即可阻止此位点被接近。因此，定位的核小体对于其附近 DNA 序列的可接近性具有正负双重效应。框 7-3 描述了确定核小体定位的一种方法。

组蛋白 N 端尾巴的修饰可改变染色质的易接近性

当组蛋白从细胞中分离出来，其 N 端尾巴通常被许多种小分子修饰（图 7-39）。尾巴中的赖氨酸经常被乙酰基或甲基基因修饰，已知至少一个精氨酸也被甲基化。同样，而丝氨酸主要受磷酸化修饰。假说认为，组蛋白尾巴的修饰构成一种“组蛋白密码”，可以被基因表达相关的蛋白及其他一些与 DNA 相互作用的蛋白识别（图 7-39、表 7-7）。按照这种假说，除了基因组序列外，与待定基因相关的核小体以及发生在这些核小体上的修饰，也能对相关基因的表达调控产生重要影响。

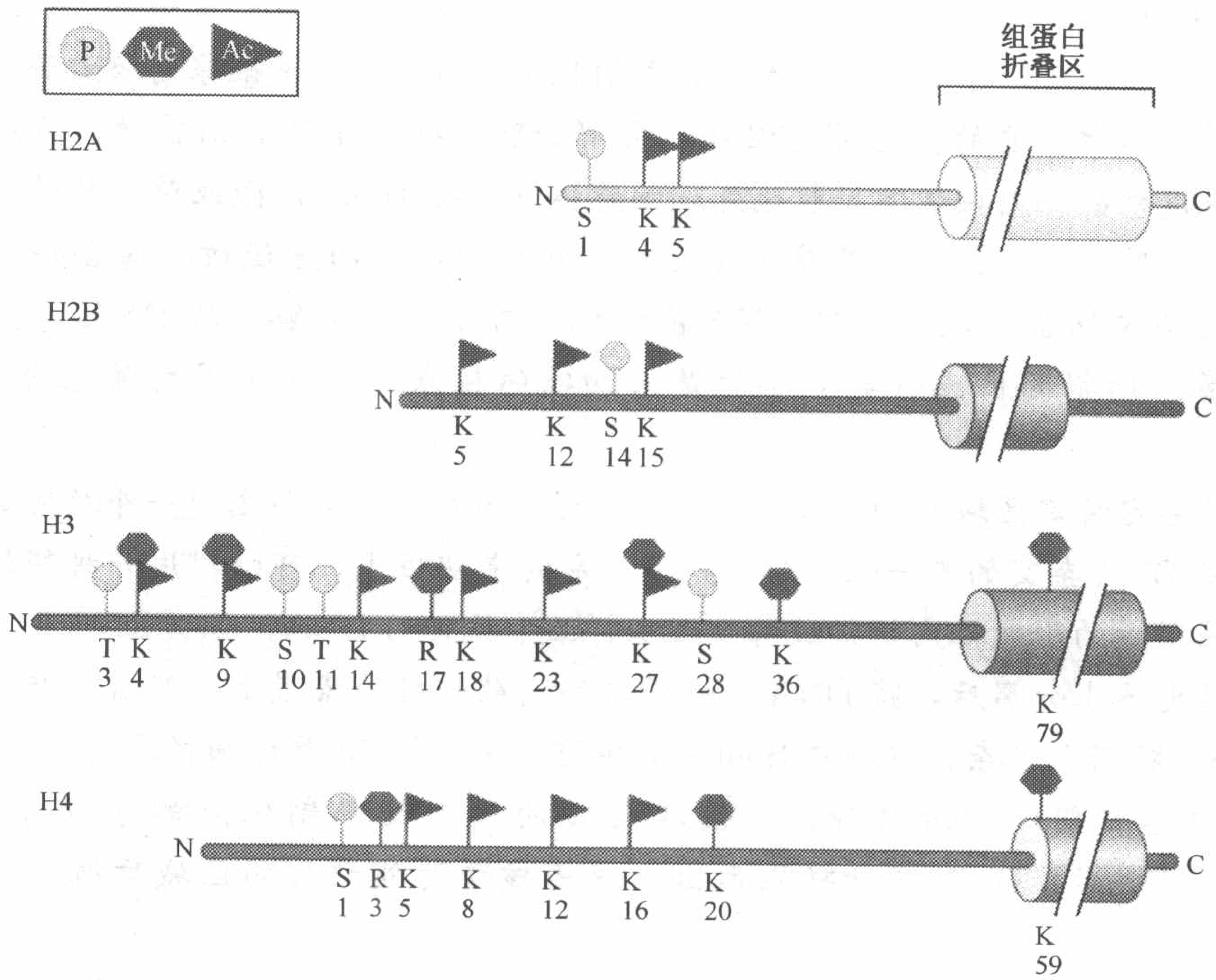


图 7-39 组蛋白 N 端尾巴的修饰改变了染色质的功能。图中显示了每个组蛋白上已知的修饰位点。大多数修饰发生在组蛋白的尾巴区域，但偶尔也发生在组蛋白折叠内部（如组蛋白 H3 的 79 位赖氨酸的甲基化）。组蛋白修饰所产生的影响依赖于修饰的类型和修饰的位点。（来源：Alberts B. et al. 2002. Molecular biology of the cell，4th edition，p. 215，Fig. 4-35. © Garland Science/Taylor & Francis LLC 和 Jenuwein and Allis. C. D. 2001. Science 293：1074-1080，Fig. 2 and 3. © AAAS，经许可）

重要的是，修饰的类型和被修饰的位置都对理解这种密码至关重要。例如：在组蛋白H4N基端第8和第16氨基酸位点上发生的赖氨酸乙酰化与基因表达的起始位置相关，而发生在第5和第12氨基酸位点上的赖氨酸乙酰化就不具有这种意义，但却是新合成的H4分子的标志，意味着相应H4分子将与DNA被整合进新的核小体中。相似地，组蛋白尾巴的甲基化也可具有不同的生物学意义，组蛋白H3第4，36或79位点上的赖氨酸甲基化与基因激活相关，而第9或第27位点上的赖氨酸甲基化则与转录抑制相关。另外，有丝分裂染色体中的高度折叠的染色质，其组蛋白H3N端通常会发生磷酸化修饰。

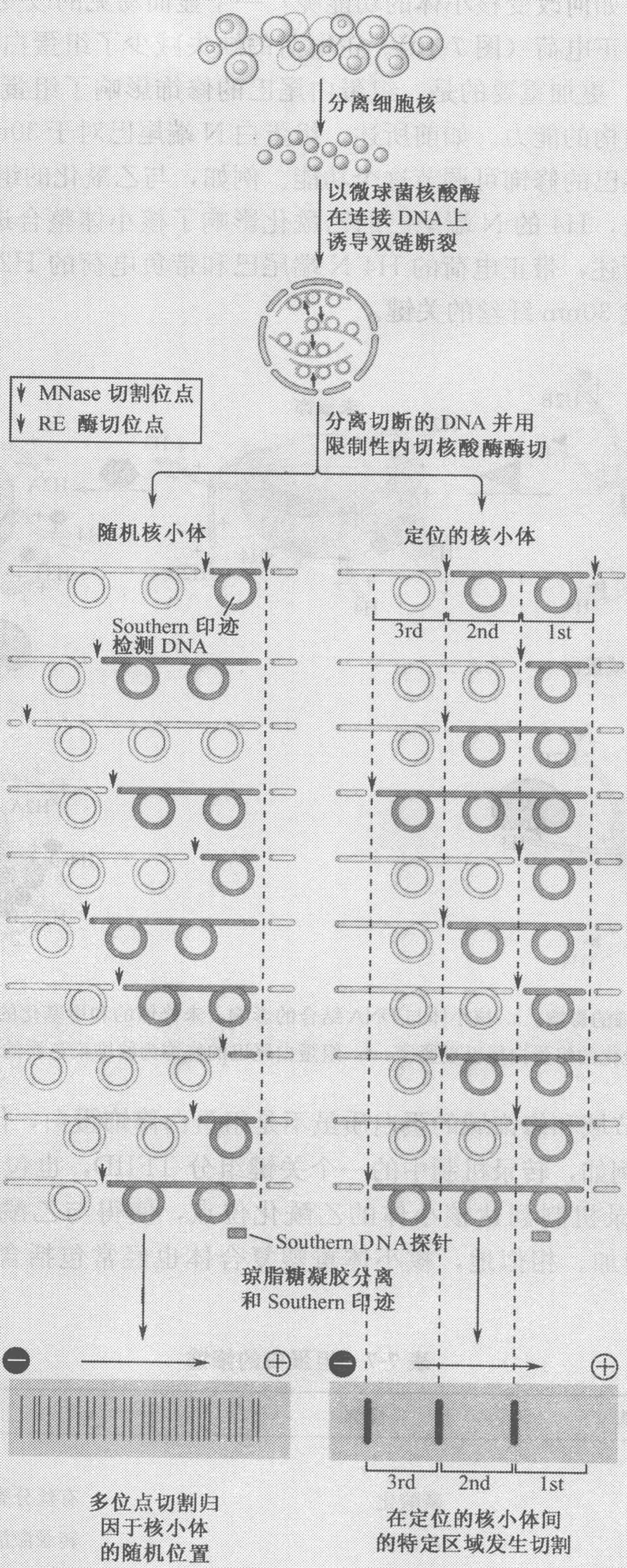
框 7-3 细胞中核小体定位的实验

核小体与相邻的重要调控序列间定位的重要性促进了在细胞中监控核小体定位方法的发展。这其中的许多方法都是利用核小体具有保护DNA免受微球菌核酸酶消化的能力。正如框7-1所描述的，微球菌核酸酶易于切断核小体间的DNA，而不是与核小体紧密结合的DNA。利用这个性质可以在一个细胞群体中定位与结合位置相同的核小体（框7-3图1）。

为了精确地定位核小体的位置，分离细胞染色质并用适量微球菌核酸酶处理使之轻微断裂是非常重要的。通常是温和地裂解细胞，得到完整的细胞核，然后再用几种不同浓度的微球菌核酸酶短暂地处理细胞核（一般1min），微球菌核酸酶很小能迅速地进入细胞核。利用不同浓度微球菌核酸酶处理的目的是使核酸酶在每个细胞中对目的区域仅切割一次。一旦DNA被消化，将细胞核裂解，从DNA上去除全部的蛋白质。切割的位点（更重要的是未切割的位点）留下了蛋白质结合DNA的记录。

为了鉴定特定区域的切割位点，有必要对所有的切割片段创建一个确定的末端位点并研究DNA杂交的专一性。为了创建固定的末端位点，可以利用一种酶切位点在目的区域附近的限制性内切核酸酶处理从样品中纯化的DNA。根据DNA的大小在琼脂糖凝胶电泳上分离后，将DNA变性并转移到硝酸纤维素膜上。然后用特定序列的标记DNA探针与之杂交（称作Southern印迹，将在第20章详细描述）。在这种情况下，DNA探针需仔细地选择，以使其与目的位点处限制位点的紧邻序列杂交。经杂交和洗脱后，DNA探针将显示出微球菌核酸酶处理目的区域后所产生片段的长度。

片段的长度是如何揭示定位的核小体的位置呢？与定位的核小体结合的DNA将不被抗微球菌核酸酶消化，而产生约160～200bp的没有被切断的DNA区域。在Southern印迹中显示为DNA条带梯度中有一个大的间隙。经常能发现有定位的核小体束存在，表现为一个相似的周期性的160～200bp片段循环，代表了切割和保护的位点。



框 7-3 图 1　细胞中核小体定位的分析。此图描述了在细胞中确定核小体定位的实验步骤。细节参阅正文。

组蛋白的修饰是如何改变核小体的功能呢？一个显而易见的改变是乙酰化和磷酸化减少了组蛋白尾巴的正电荷（图 7-40）。正电荷的缺失减少了组蛋白尾巴对带负电荷的DNA 骨架的亲和性。更加重要的是，组蛋白尾巴的修饰影响了组蛋白束形成更为抑制状态的高级染色质结构的能力。如前所述，组蛋白 N 端尾巴对于 30nm 纤丝的形成是必需的，而对组蛋白尾巴的修饰可调节这个功能。例如，与乙酰化的组蛋白和基因组的表达区结合相一致的是，H4 的 N 端尾巴的乙酰化影响了核小体整合进抑制状态的 30nm 纤丝的能力。如前所述，带正电荷的 H4 N 端尾巴和带负电荷的 H2A 组蛋白折叠区表面的相互作用是形成 30nm 纤丝的关键。

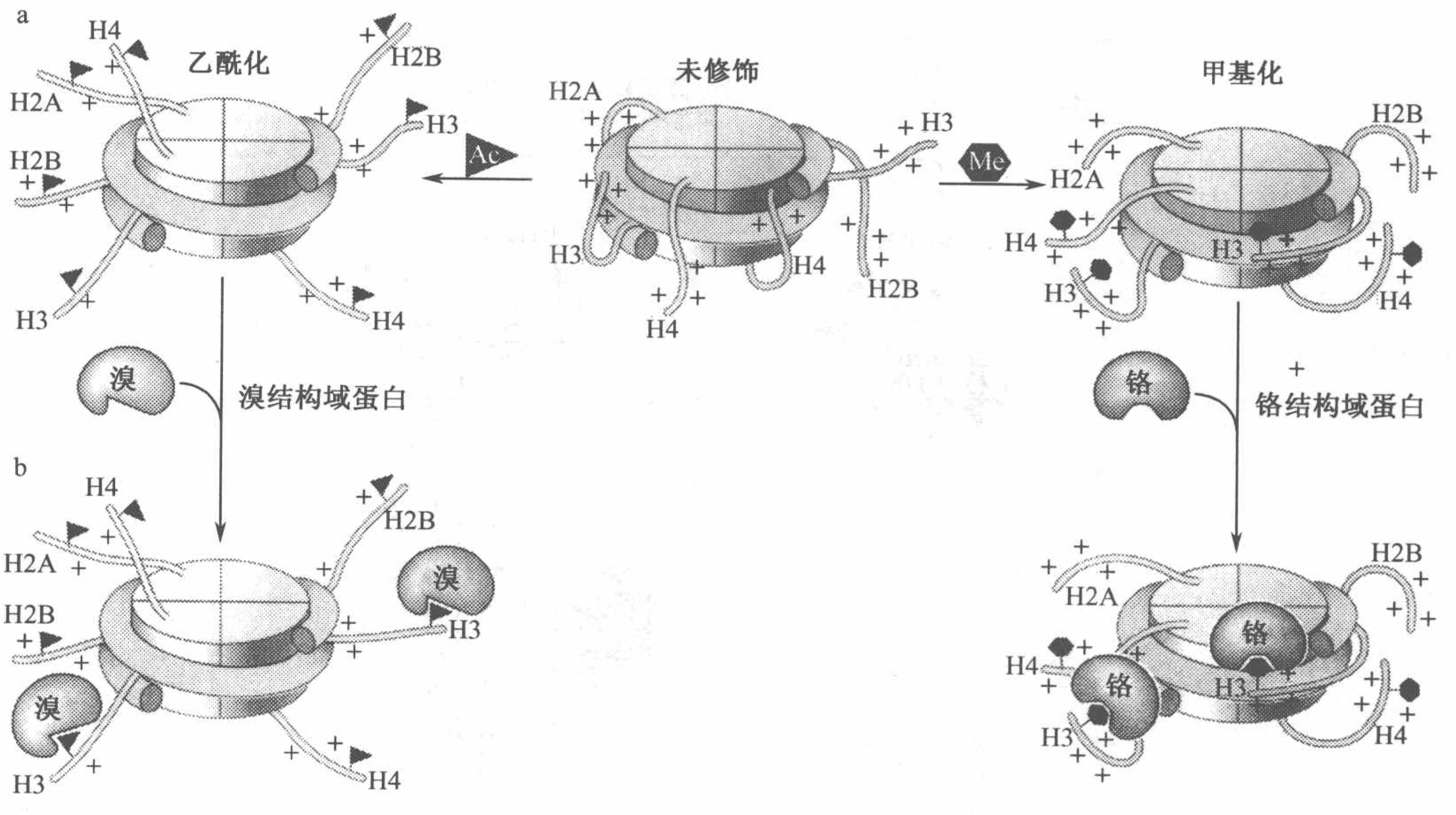


图 7-40 组蛋白尾巴修饰的影响。a. 核小体与 DNA 结合的影响。未修饰的和甲基化的组蛋白尾巴被认为与核小体 DNA 的结合比乙酰化的组蛋白尾巴更紧密。b. 组蛋白尾巴的修饰为修饰染色质的酶提供了结合位点。

其他含同源调节域和修饰域的蛋白质虽不是组蛋白修饰蛋白，但参与转录调控或异源染色质的形成。例如，转录机制中的一个关键组分 TFIID，也包含一个同源调节域。这个结构域指导转录机制到达核小体的乙酰化位点，使得与乙酰化的核小体结合的DNA 的转录活性增加。相似地，核小体重塑复合体也经常包括含同源调节域的亚基（表 7-7）。

表 7-7 组蛋白的修饰

组蛋白亚基的残基	修饰	效应
H2A		
丝氨酸 1	磷酸化	有丝分裂，转录抑制
赖氨酸 4	乙碳化	转录激活
赖氨酸 5	乙酰化	转录激活
赖氨酸 7	乙酰化	转录激活

续表

组蛋白亚基的残基	修饰	效应
H2B		
赖氨酸 5	乙酰化	转录激活
赖氨酸 12	乙酰化	转录激活
丝氨酸 14	磷酸化	细胞凋亡
赖氨酸 15	乙酰化	转录激活
H3		
苏氨酸 3	磷酸化	有丝分裂
赖氨酸 4	乙酰化、甲基化	乙酰化：转录激活 甲基化：活性化染色质
赖氨酸 9	乙酰化、甲基化	乙酰化：转录激活 甲基化：转录抑制
丝氨酸 10	磷酸化	转录激活
苏氨酸 11	磷酸化	有丝分裂
赖氨酸 14	乙酰化	转录激活-延伸
精氨酸 17	甲基化	转录激活
赖氨酸 18	乙酰化	转录激活、DNA 修补
赖氨酸 23	乙酰化	转录激活、DNA 修补
赖氨酸 27	甲基化、乙酰化	转录沉默
丝氨酸 28	磷酸化	有丝分裂
赖氨酸 36	甲基化	转录延伸
赖氨酸 56	乙酰化	DNA 修补
赖氨酸 79	甲基化	转录延伸
H4		
丝氨酸 1	磷酸化	有丝分裂
精氨酸 3	甲基化	转录激活
赖氨酸 5	乙酰化	组蛋白沉淀
赖氨酸 8	乙酰化	转录激活
赖氨酸 12	乙酰化	组蛋白沉淀、端粒沉默
赖氨酸 16	乙酰化	转录激活、DNA 修补
赖氨酸 20	甲基化	转录沉默

a. 据 Peterson C. L. and Lariel M. A. 2004. Curr Biol. 14，R546-R551 修改

b. 表示脊椎动物的已知修饰。大多数修饰也见于其他生物，尽管有时不同生物的组蛋白结构有所不同，因而不同修饰的位置编号可能不动。

核小体重塑和修饰复合物的蛋白结构域识别经修饰的组蛋白

经修饰的组蛋白尾巴能起聚集特异性蛋白到染色质上的作用（图 7-40b）。溴结构

域（bromodomains）铬结构域（chromodomains），TUDOR 结构域和 PHD 指纹（植物同源结构域，homodomain）能识别经修饰的形式的组蛋白尾巴，那些含溴结构域的蛋白能与乙酰化的组蛋白尾巴相互作用，而那些含有铬结构域-TUDOR 结构域和 PHD 指纹的蛋白质能与甲基化的组蛋白相互作用。另一个称为 SANT 结构域的蛋白结构域的特点正好相反，那些含 SANT 结构域的蛋白质优先与未经修饰的组蛋白尾巴相互作用，与这些蛋白结构域对于解释组蛋白修饰的重要性一致，很多情况下含这些结构域的蛋白质特异性地识别很多组蛋白修饰可能位点中的一个修饰形式。例如，HP1 蛋白含有一个铬结构域，它只能与 9 位赖氨酸甲基化了的 H3 组蛋白结合，而不与任何其他位置被甲基化的组蛋白结合。有趣的是，还有很多蛋白质也含有这样的一个结构域，说明这些蛋白质是特化了，专门识别多种修饰的组蛋白。例如，一些蛋白质含有 PHD 指纹，紧靠溴结构域的组蛋白 H3 的甲基化了的 4 粒氨酸，而这一溴结构域能识别乙酰化的赖氨酸。

这些能识别经修饰的组蛋白的结构域是如何改变相关核小体的功能的？一个重要的方式是经修饰的组蛋白能将很多酶聚集起来，进一步修饰邻近的核小体。例如，很多这些修饰组蛋白尾巴的酶（称为组蛋白乙酰基转移酶，hisfone acetyltransferase）含有溴结构域，这些溴结构域能识别它们产生的同样的组蛋白修饰（表 7-8），在这种情况下，溴结构域能进一步修饰已经乙酰化的区域，有助于乙酰化组蛋白的维持和继续（见下面的讨论）。

经修饰的组蛋白也能聚集作用于染色质的其他蛋白。很多核小体重塑复合物的一个或几个亚基都有这些识别修饰了的组蛋白的结构域（见表 7-6），使修饰了的组蛋白能够这些酶。很多与转录调控有关的蛋白也含有这些结构域。例如，TF11D 转录机器的关键组分就含有一个溴结构域，这个结构域将转录机器引导到核小体乙酰化位置上。这是组蛋白乙酰化的另一个途径，这一乙酰化能增加相关 DNA 的转录活性。能识别组蛋白甲基化位置的铬结构域可以与转录被抑制的基因结合，这个铬结构域在好多蛋白质都已发现。这些蛋白质对于异染色质的形成非常重要，包括 HP1 蛋白和“多梳（polycomb)”蛋白。

特定的酶负责组蛋白的修饰

我们刚才已经讲述了组蛋白的修饰是动态的，且是由特定的酶催化的。组蛋白乙酰转移酶催化乙酰基团添加到组蛋白上，而组蛋白去乙酰化酶可去除这种修饰。相似地，组蛋白甲基转移酶将甲基加到组蛋白上组蛋白去甲基酶则将甲基去除。许多不同的组蛋白乙酰转移酶和去乙酰化酶已经被鉴定，可通过它们对不同的组蛋白或是同一组蛋白不同赖氨酸的识别能力上来进行区分。组蛋白甲基转移酶和去甲基化酶是有更好的特异性，通常可以精确识别组蛋白上众多赖氨酸或精氨酶中的仅仅某一个（表 7-8）。由于不同的修饰对核小体的功能有不同的影响，因此不同的组蛋白乙酰转移酶或甲基转移酶对核小体的修饰会对染色质的结构和 DNA 的很多功能进行调控（图 7-39）。

表 7-8 组蛋白修饰酶

组蛋白乙酰转移酶复合体				
类型	亚基数目	催化亚基	组蛋白结合域	靶组蛋白
SAGA	15	Gcn5	铬功能域、溴功能域	H3 和 H2B
PCAF	11	PCAF	溴功能域	H3 和 H4
NuA3	3	Sas3	PHD 指纹	H3
NuA4	6	Esa1	铬功能域、SANT 域、PHD 指纹	H4 和 H2A
P300/CBP	1	P300/CBP	溴功能域、PHD	H2A、H2B、H3 和 H4

组蛋白去乙酰化酶复合体			
类型	亚基数目	催化亚基	同源调节域/修饰域
NuRD	9	HDAC1/HDAC2	溴功能域、PHD 指纹
SIR2 复合物	3	Sir2	无
Rpd3 大亚基	12	Rpd3	PHD 指纹
Rpd3 小亚基	5	RpD3	溴功能域，PHD 指纹

组蛋白甲基转移酶		
名称	组蛋白结合域	靶组蛋白
SET1	无	H3（赖氨酸 4）
SUV39/CLR4	溴功能域	H3（赖氨酸 9）
SET2	无	H3（赖氨酸 36）
DOT1	无	H3（赖氨酸 79）
PRMT	无	H3（精氨酸 3）
SET9/SUV4-20	无	H4（赖氨酸 20）

组蛋白去甲基化酶		
名称	组蛋白结合域	靶甲基化组蛋白
LSD1	PHD 指纹、SANT 域	H3（赖氨酸 4）
JHDM1	PHD 指纹	H3（赖氨酸 36）
JHDM3	PHD 指纹，TUDOR 域	H3（赖氨酸 9，36）

与核小体重塑复合体一样，这些修饰酶是大的多聚蛋白复合体的一部分。其他的亚基在引导这些酶到特定的 DNA 区域上起着重要作用。与核小体重塑复合体相似，这些复合体可以和与 DNA 结合的转录因子或直接与修饰的核小体发生相互作用。引导这些酶到特定的 DNA 区域是沿着染色质能够观察到特定的组蛋白修饰模式的原因，并且是调控真核染色体中基因表达水平的主要机制（见第 17 章）。

核小体的修饰和重塑共同增加 DNA 的易接近性

N 端尾巴的修饰和核小体重塑的共同作用能显著地改变 DNA 的易接近性。正如我们将在第 12 章和第 17 章中所学到的，参与这些修饰的蛋白质复合体经常集合到转录活性位点上。尽管它们发挥作用的顺序不总是相同，但累积起来的效果是使 DNA 的易接近性在特定的部位产生显著的改变。N 端尾巴的修饰能减少核小体束形成抑制性结构

的能力，产生能集合包括核小体重塑因子在内的其他蛋白质的位点。随后的核小体重塑能进一步增加核小体 DNA 的易接近性，允许 DNA 结合蛋白接近它们的结合位点。与适当的 DNA 结合蛋白或 DNA 序列相结合，这些改变能导致 DNA 特定位点处的核小体的定位或释放（图 7-41）。

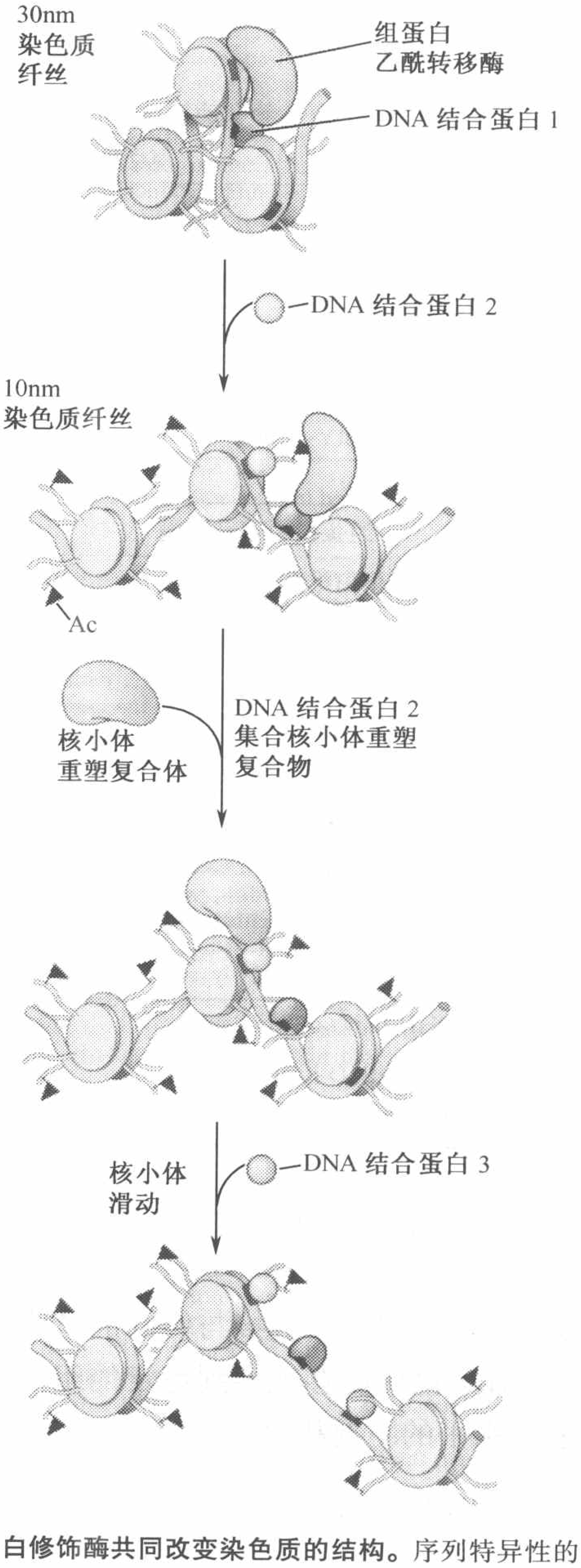


图 7-41 染色质重塑复合体和组蛋白修饰酶共同改变染色质的结构。序列特异性的 DNA 结合蛋白通常引导这些酶到染色体的特定区域。在这个图中，DNA 结合蛋白（蓝色）首先引导一个组蛋白乙酰转移酶来修饰邻近的核小体，使局部的 30nm 染色质纤丝转变为较易接近的 10mm 纤丝，增加相关的 DNA 的易接近性。这导致了第二个 DNA 结合蛋白（橙色）的结合，它又导入了一个核小体重塑复合物，核小体重塑复合物的定位有助于邻近核小体的滑动，暴露了第三个结合蛋白（绿色）的结合位置，如转录起始位点的 TATA 结合蛋白质的结合位点。尽管我们在这里描述的是先组蛋白乙酰化复合体结合，后组染小体重塑复合体的顺序，但两种顺序都可以观察到并具有相同的效果。事实上，不同的组蛋白修饰酶，会导致形成更浓缩的、不易接近的染色质。

核小体的组装

DNA 复制后核小体组装即开始

一条染色体的复制需要 DNA 的复制和子代 DNA 分子上相关蛋白质的重新组装。后一过程与 DNA 复制紧密相连，以确保新复制的 DNA 分子迅速组装进核小体。在第 8 章中我们将详细讨论 DNA 复制的机制。这里我们只讨论 DNA 复制后指导核小体组装的机制（见互动漫画 7-2）。

虽然 DNA 的复制要求核小体的部分解体，但是 DNA 在一系列有序的过程中很快被重新组装。如前所述，在 DNA 上核小体组装的第一步是结合一个 H3 和 H4 四聚体。一旦四聚体结合后，两个 H2A · H2B 二聚体结合形成最终的核小体。据推测，在高级染色质组装的过程中，H1 最后加入这个复合体。

为了复制一条染色体，子代染色体上的核小体至少半数是新合成的。是不是将全部的原有组蛋白丢失而仅用新合成的组蛋白来组装成核小体？如果不是这样，原有的组蛋白是如何在两个子代染色体间分配？考虑到组蛋白的修饰会对染色质的易接近性有影响，因此原有组蛋白的命运就成为一个特别重要的问题。如果原有的组蛋白完全丢失，那么染色体的复制将删除对先前修饰的核小体的任何"记忆"。相反，如果原有的组蛋白保留在一条染色体上，那么这条染色体相对于另一拷贝的染色体将有一套不同的修饰。

在实验中采用不同的标记物标记原有的和新的组蛋白，发现原有的组蛋白存在于两条子代染色体中（图 7-42）。然而，这种混合不是完全随机的。H3 · H4 四聚体和 H2A · H2B 二聚体或是全由新的组蛋白或者全由旧的组蛋白所组成。因此，当复制叉经过时，核小体解体为亚组装部件。H3 · H4 四聚体似乎仍随机地与两个子代双螺旋之一结合，并不从 DNA 上释放而成为游离的组蛋白成分。相反，H2A · H2B 二聚体被释放且进入局部的环境，参与新的核小体的组装。

在染色体复制过程中原有组蛋白分布的遗传性为精确地按父本模式进行组蛋白修饰提供了一个机制。通过这个机制，无论原有的组蛋白在哪个子代染色体上，都倾向于重新结合到靠近它们在父代染色体上的位置上（图 7-43）。这种修饰组蛋白定位的遗传性为每个子代染色体上相似的位置提供了数量有限的修饰。这些组蛋白修饰具有引导修饰酶的能力，可以在邻近的核小体上进行相似的修饰（见上面对组蛋白结合域的讨论）。这为在 DNA 复制发生后维持相似的修饰状态提供了一个简单的机制（图 7-43）。这些机制可能对染色质状态的世代遗传具有重要作用。考虑到组蛋白修饰在基因表达调控以及其他一些 DNA 的相互作用中的重要性，保持这些修饰状态对维持细胞特性以及细胞中 DNA 的复制与分离都是很关键的。

核小体的组装需要组蛋白"伴侣"

核小体的组装不是一个自发的过程。早期的研究发现简单的将纯化的组蛋白添加到 DNA 中，仅有很少的甚至没有核小体形成。相反，大多数组蛋白聚集形成没有活性的形

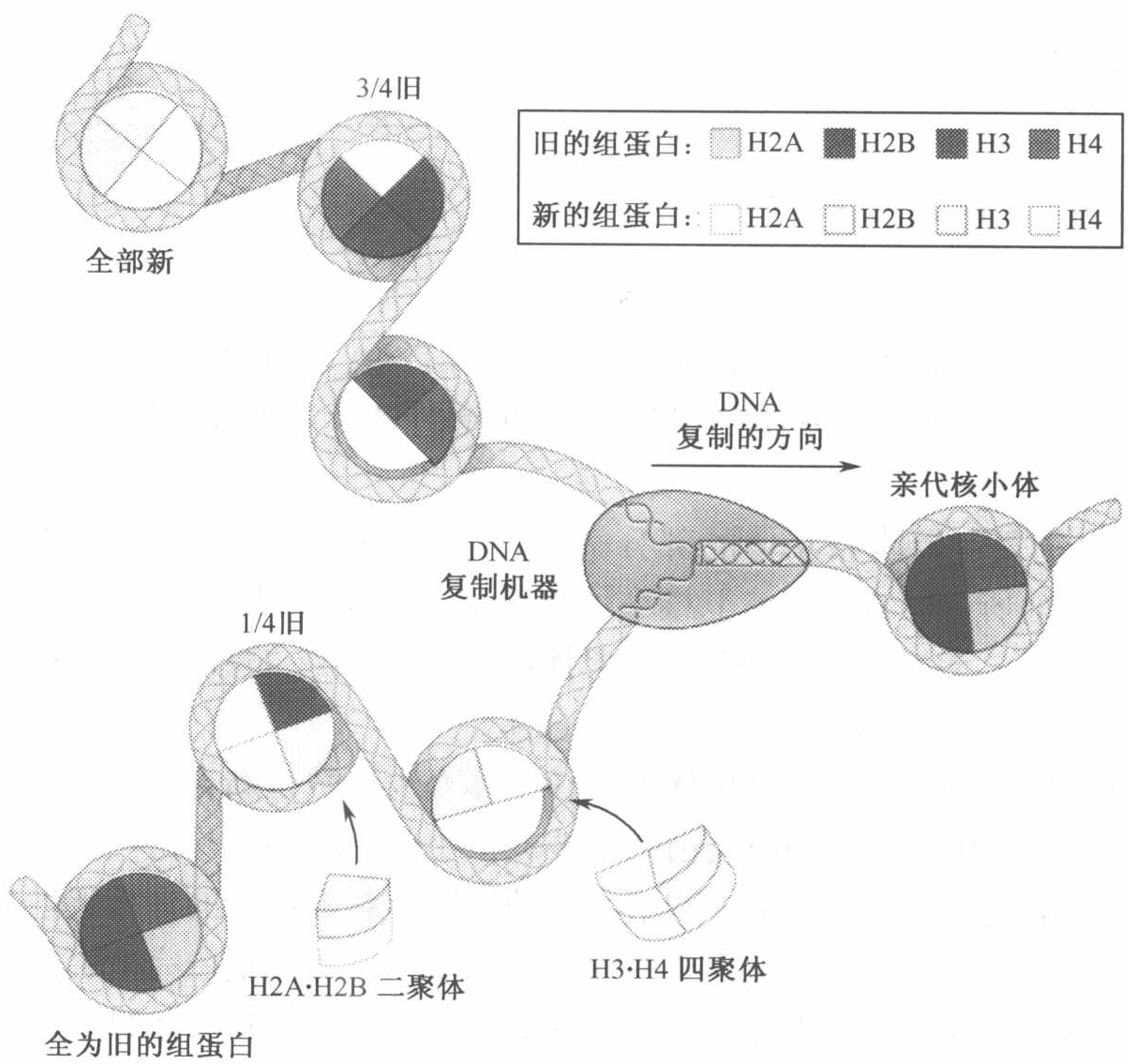


图 7-42　DNA 复制后组蛋白的遗传。染色体复制时，与父本染色体结合的组蛋白具有不同的分配形式。组蛋白 H3 · H4 四聚体随机地转移到两条子链中的任一条上，但不解离为游离状态的 H3 · H4 四聚体。新合成的 H3 · H4 四聚体在没有继承亲代四聚体的链上形成核小体的基础；相反，H2A 和 H2B 二聚体被释放进入游离的可溶环境，并与新合成的 H2A 和 H2B 竞争与 H3 · H4 的结合。一般来说，这种分配类型的后果是新合成的 DNA 上的第二个 H3 · H4 四聚体将来自亲代染色体。这些四聚体保留了全部父代核小体上的修饰；H2A 和 H2B 二聚体则更可能是新合成的。

式。为了正确的组装核小体，有必要将盐浓度提高到非常高的水平（大于 1mol/L NaCl），然后再在数小时内缓慢降低浓度。尽管提高盐浓度有助于核小体组装的离体研究（如前面讨论的对核小体结构的研究），但在体内核小体的组装并不涉及盐浓度的升高。

在生理盐浓度下对核小体组装的研究鉴定出一些能指导组蛋白在 DNA 上组装的因子。这些因子是带负电荷的蛋白质，与 H3 · H4 四聚体或 H2A · H2B 二聚体形成复合体（表 7-9），并护送它们到核小体组装的位置。由于这些因子可以避免组蛋白与 DNA 发生无效的相互作用，因此被称作组蛋白伴侣（histone chaperone）（图 7-44）。

表 7-9　组蛋白伴侣的特征

名称	亚基数目	结合组蛋白	与滑动夹是否相互作用
CAF-1	4	H3 · H4	是
H1RA	4	H3 · H4	否
RCAF	1	H3 · H4	否
NAP-1	1	H2A · H2B	否

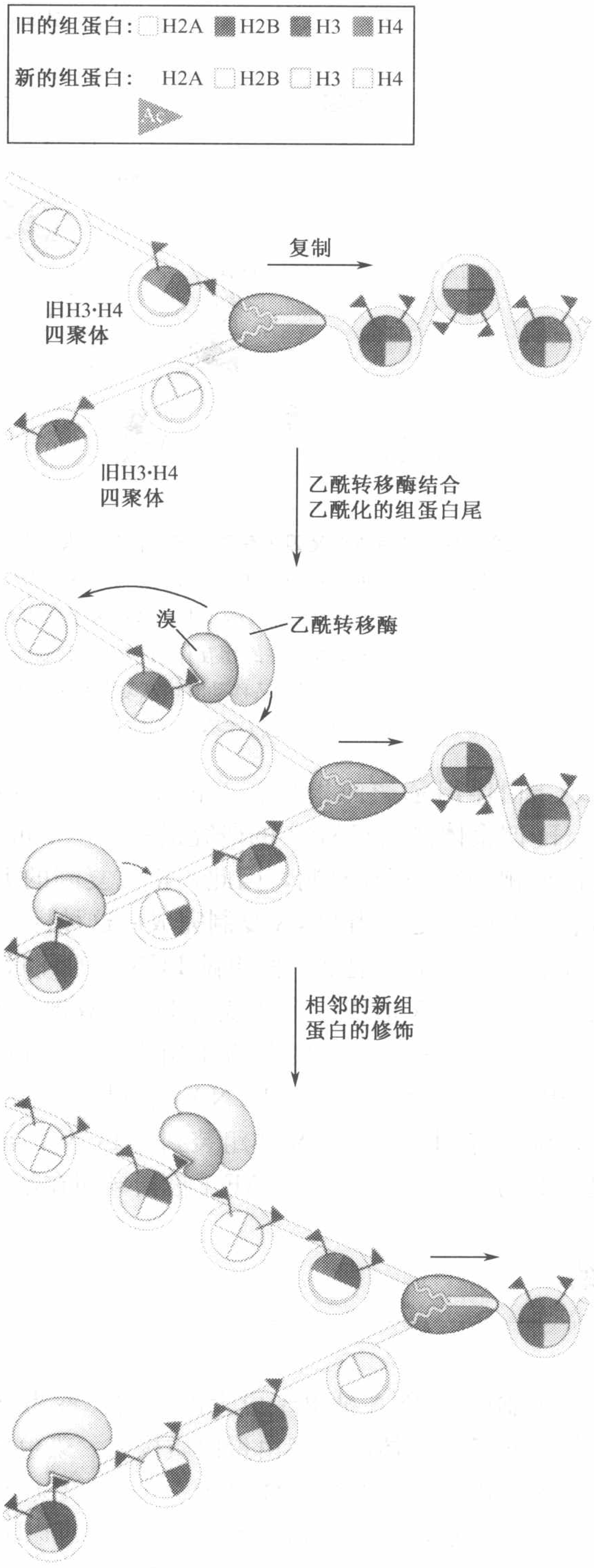


图 7-43 亲代 H3 · H4 四聚体的遗传有助于染色质状态的遗传。染色体复制后，父代 H3 · H4 四聚体的分配导致子代染色体获得了与亲代相同的修饰。这些修饰具有导入相同修饰酶的能力，有助于将相同的修饰状态正确传递到两个子代染色体。为简便起见，图中显示在组蛋白核心区的乙酰化。事实上，这种修饰一般发生在组蛋白的N端尾。

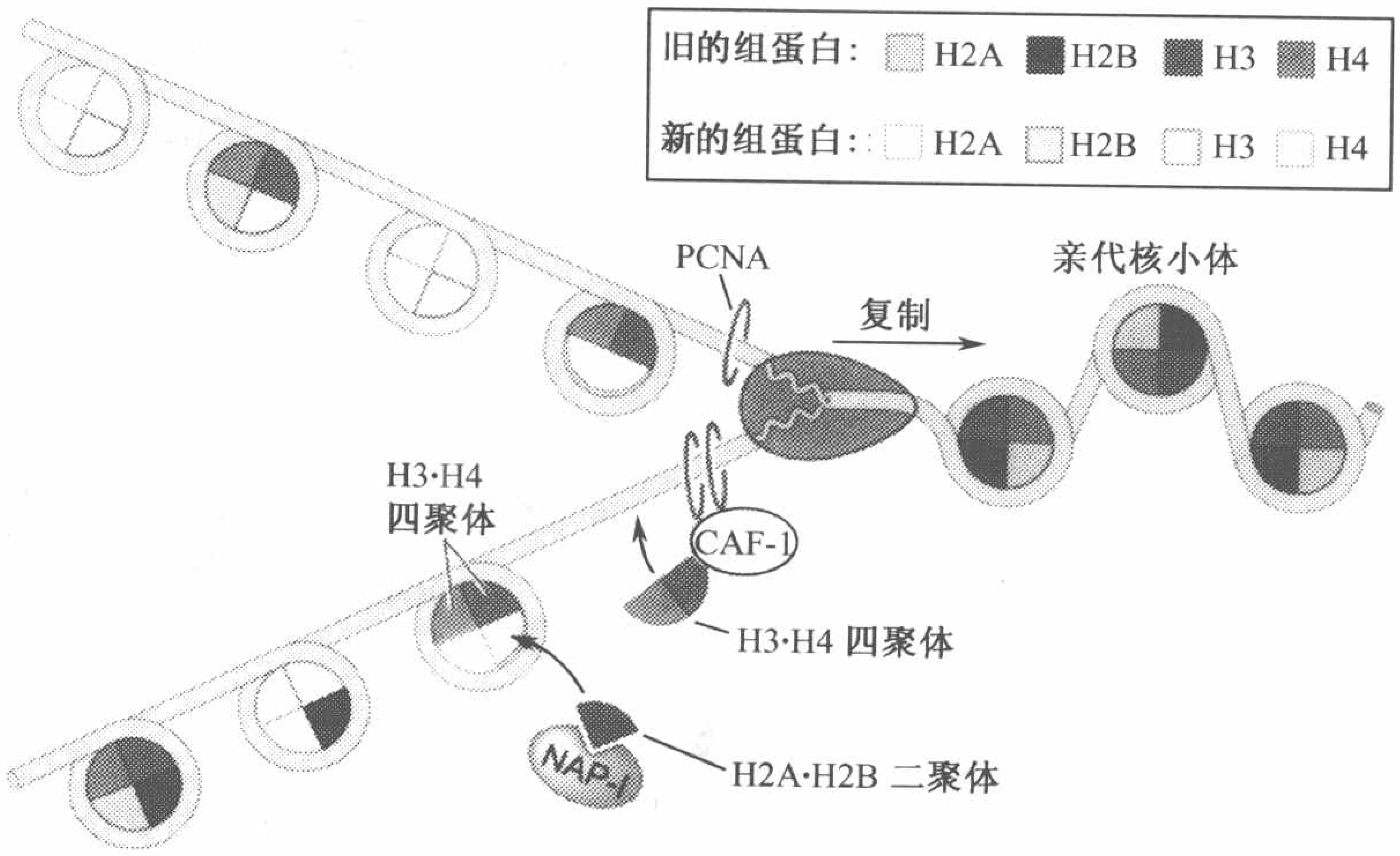


图 7-44 染色质组装因子有助于核小体的组装。复制叉经过后，染色质组装因子伴随游离的 H3 · H4 四聚体（如 CAF-1）和 H2A · H2B 二聚体（NAP-1）到新复制 DNA 的位点。一旦到了新复制的 DNA 上，这些因子即将其结合的组蛋白转移给 DNA。CAF-1 因子通过与 DNA 滑动夹的相互作用被引导到新复制的 DNA 上。这些环状的辅助复制因子环绕 DNA，当复制叉移动时可从复制机器上释放。有关更为详细的 DNA 滑动夹及其在 DNA 复制中的功能将在第 8 章讲述。

组蛋白伴侣是如何指导在新的 DNA 合成位置进行核小体的组装呢？对组蛋白 H3 · H4 四聚体伴侣 CAF-1 的研究揭示了一个可能的答案。由 CAF-1 指导的核小体的组装需要靶 DNA 正在复制，因此，正在复制的 DNA 以某种方式被标记，用于核小体的组装。有趣的是，当 DNA 复制完成后这种标记就逐渐消失。对依赖于 CAF-1 的组装的研究表明，这个标记是一种叫做 PCNA 的环状滑动夹蛋白。正如将在第 8 章详细讨论的一样，PCNA 形成一个环绕着 DNA 双螺旋的环，在 DNA 合成过程中将 DNA 聚合酶固定在 DNA 上。当聚合酶作用完成后，PCNA 由 DNA 聚合酶上释放，但仍与 DNA 相连。在这种情况下，PCNA 能与其他蛋白质发生作用。CAF-1 与释放的 PCNA 结合，优先将 H3 · H4 四聚体组装到与 PCNA 结合的 DNA 上。因此，通过与 DNA 复制机制上的一个成分结合，CAF-1 被指导而在新复制的 DNA 位点组装核小体。

小结

在细胞中，DNA 被组织成大的结构，称为染色体。尽管 DNA 是染色体形成的基础，但是每条染色体的一半由蛋白质构成。染色体可能是线性的或环状的；但每个细胞有特定数目的染色体和染色体组成。我们现在已经知道了几千种生物体完整的基因组序列。这些序列显示每个生物染色体上的 DNA 以不同的效率来编码蛋白质。简单的生物使用绝大多数的 DNA 编码蛋白质；然而，较复杂的生物仅用一小部分 DNA 编码蛋白质或 RNA。

当细胞分裂时，细胞必须精确维持其染色体的组成。每个染色体必须含有在细胞分裂时指导染色体维持的 DNA 元件。所有的染色体必须有一个或多个复制起始位点。在真核细胞中，着丝粒在染色体分离过程中起着至关重要的作用，而端粒帮助保护和复制线性染色体的末端。在细胞分裂过程中，真核细胞精细地将染色体的复制和分离过程分开。染色体的分离有两种方式。在有丝分裂过程中，一种高度特化的机制确保每个复制的染色体的一个拷贝传递到每个子代细胞中；在减数分裂过程中，额外的一轮染色体分裂（没有 DNA 复制）进一步减少了子代细胞中染色体的数目。

真核 DNA 及其相关蛋白质的结合域染色质。染色质的基本单位是核小体，由各两拷贝的核心组蛋白（H2A、H2B、H3、H4）和大约 147bp 的 DNA 构成。这个蛋白质-DNA 复合物在细胞中有两个重要的功能：压缩 DNA 以适应细胞核的大小和限制 DNA 的易接近性。细胞广泛地利用后一种功能来调控许多不同的 DNA 功能，包括基因表达。

核小体的原子结构显示 DNA 盘绕在圆盘状的组蛋白核心外大约 1.7 圈。DNA 和组蛋白的相互作用非常广泛，但都没有碱基的特异性。这些相互作用的特点解释了 DNA 绕组蛋白八聚体的弯曲和几乎所有 DNA 序列最终组装进核小体的能力。这个结构也揭示了组蛋白 N 端尾巴的位置及其指导 DNA 进出核小体的路径。

一旦 DNA 组装成核小体，它有能力形成更复杂的结构来进一步压缩 DNA。第五种组蛋白 H1 有助于这个过程的完成。H1 通过与结合在核小体上的 DNA 结合，使得 DNA 更紧密地盘绕在八聚体上。一种更为压缩的染色质形式，30nm 纤丝，是由结合有 H1 的核小体束有序排列构成。这个结构比 DNA 仅仅形成核小体更具有抑制性。已有的证据表明，形成这种结构的 DNA 与参与 DNA 转录的酶或蛋白质的易接近性显著降低。

在核小体中 DNA 与组蛋白的结合是动态的，允许 DNA 结合蛋白间歇地靠近 DNA。核小体重塑复合物通过增加核小体的可移动性来增加核小体上的 DNA 的易接近性。可观察到 2 种运动方式：组蛋白八聚体沿 DNA 的滑动以及组蛋白八聚体从一个 DNA 完整地转移到另一个 DNA 分子。这些复合体定位于基因组的特定区域，以便于染色质易接近性的改变。某些核小体被限定在基因组的固定位置上，称作“定位”。核小体的定位由 DNA 结合蛋白或特殊的 DNA 序列所指导。

组蛋白 N 端尾巴的修饰也可以改变染色质的易接近性。修饰的类型包括赖氨酸的乙酰化和甲基化及丝氨酸的磷酸化。N 端尾巴的乙酰化经常发生于基因表达的活跃区。这些修饰既改变了核小体自身的特性，也作为蛋白质的结合位点，由此影响染色质的易接近性。这些修饰也会招集具有相同修饰作用的酶，对相邻核小体进行相似的修饰。这可能使染色体复制时被修饰的核小体/染色质的区域得以稳定地传递。

DNA 复制后，核小体立即被组装，DNA 呈未被包装状态的存在时间极短。这涉及特定的组蛋白伴侣的功能，这些组蛋白伴侣护送 H3 · H4 四聚体和 H2A · H2B 二聚体到达复制叉。在 DNA 的复制过程中，核小体暂时解体。组蛋白 H3 · H4 四聚体和 H2A · H2B 二聚体被随机地分配到任一子代分子中。一般来说，每个新的 DNA 分子中有一半旧的和一半新的组蛋白。这样，两组染色体都承继了修饰的组蛋白，这些修饰

的组蛋白然后作为“种子”，使邻近的组蛋白发生相似的修饰。

（娄晓敏　译　胡松年　刘韧　杨焕明　校）

参考文献

书

Allis C.D., Jenuwein T., Reinberg D., and Caparros M.-L., eds. 2007. *Epigenetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Brown T.A. 2007. *Genomes 3*, 2nd ed. Garland Science, New York.

Morgan D.O. 2007. *The cell cycle: Principles of control*. New Science Press Ltd., London.

染色体

Bendich A.J. and Drlica K. 2000. Prokaryotic and eukaryotic chromosomes: What's the difference? *Bioessays* **22:** 481–486.

Thanbichler M., Wang S.C., and Shapiro L. 2005. The bacterial nucleoid: A highly organized and dynamic structure. *J. Cell Biochem.* **96:** 506–521.

核小体

Annunziato A.T. and Hansen J.C. 2000. Role of histone acetylation in the assembly and modulation of chromatin structures. *Gene Expr.* **9:** 37–61.

Belmont A.S., Dietzel S., Nye A.C., Strukov Y.G., and Tumbar T. 1999. Large-scale chromatin structure and function. *Curr. Opin. Cell Biol.* **11:** 307–311.

de la Cruz X., Lois S., Sánchez-Molina S., and Martinez-Balbás M.A. 2005. Do protein motifs read the histone code? *Bioessays* **27:** 164–175.

Eberharter A. and Becker P.B. 2002. Histone acetylation: A switch between repressive and permissive chromatin. *EMBO Rep.* **31:** 224–229.

Gregory P.D., Wagner K., and Horz W. 2001. Histone acetylation and chromatin remodeling. *Exp. Cell. Res.* **265:** 195–202.

Hayes J.J. and Hansen J.C. 2001. Nucleosomes and the chromatin fiber. *Curr. Opin. Genet. Dev.* **11:** 124–129.

Jenuwein T. and Allis C.D. 2001. Translating the histone code. *Science* **293:** 1074–1080.

Klose R.J., Kallin E.M., and Zhang Y. 2006. JmjC-domain-containing proteins and histone demethylation. *Nat. Rev. Genet.* **7:** 715–727.

Luger K. and Richmond T.J. 1998. DNA binding within the nucleosome core. *Curr. Opin. Struct. Biol.* **8:** 33–40.

———. 1998. The histone tails of the nucleosome. *Curr. Opin. Genet. Dev.* **8:** 140–146.

Luger K., Madev A.W., and Richmond R.K. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389:** 251–260.

Margueron R., Trojer P., and Reinberg D. 2005. The key to development: Interpreting the histone code? *Curr. Opin. Genet. Dev.* **15:** 163–176.

Martin C. and Zhang Y. 2005. The diverse functions of histone lysine methylation. *Nat. Rev.* **6:** 838–849.

Narliker G.J., Fan H.-Y., and Kingston R.E. 2002. Cooperation between complexes that regulate chromatin structure and transcription. *Cell* **108:** 475–487.

Robinson P.J.J. and Rhodes D. 2006. Structure of the '30 nm' chromatin fibre: A key role for the linker histone. *Curr. Opin. Struct. Biol.* **16:** 336–343.

Roth S.Y., Denu J.M., and Allis D. 2001. Histone acetyl transferases. *Annu. Rev. Biochem.* **70:** 81–120.

Saha A., Wittmeyer J., and Cairns B.R. 2006. Chromatin remodelling: The industrial revolution of DNA around histones. *Nat. Rev. Mol. Cell. Biol.* **7:** 437–447.

Thiriet C. and Hayes J.J. 2005. Chromatin in need of a fix: Phosphorylation of H2AX connects chromatin to DNA repair. *Mol. Cell* **18:** 617–622.

Thomas J.O. 1999. Histone H1: Location and role. *Curr. Opin. Cell. Biol.* **11:** 312–317.

Woodcock C.L. and Dimitrov S. 2001. Higher-order structure of chromatin and chromosomes. *Curr. Opin. Genet. Dev.* **11:** 130–135.

第8章 DNA的复制

本章概要

- DNA 合成的化学基础
- DNA 聚合酶的作用机制
- 复制叉
- DNA 聚合酶的特化
- 复制叉上的 DNA 合成
- DNA 复制的起始
- 结合和解旋：起始子蛋白对复制起始位点的选择和激活
- 结束复制

当 DNA 双螺旋被揭示的时候，最令生物学家振奋的是 DNA 多聚核苷酸互绕链上碱基之间的互补性。似乎很难想像这种互补性的结构会不被用作 DNA 复制的基础。事实上，正是 DNA 的结构所揭示出的自身互补性质，才最终使生物学家们接受了 Oswald T. Avery 的论断，即遗传信息的载体是 DNA，而不是某种形式的蛋白质（第 2 章）。

当讨论模板如何作用的时候，我们强调两个具有相同表面的基团不会相互吸引（第 6 章）；相反，具有相反形状或电荷基团之间的吸引则容易得多。因此，无需任何结构上的认识，也能猜测到像基因这样复杂的分子不可能直接被复制；相反，复制会涉及形成具有形状上互补的分子，并作为复制的模板。因此，在对蛋白质或核酸结构取得详尽的认识之前，一些遗传学家曾怀疑 DNA 是否作为某一特定蛋白质的模板，而此蛋白质反过来也是相应 DNA 的模板。

但是自 DNA 自身互补的性质被揭示后，关于蛋白质模板可能参与 DNA 复制的观点很快就被否定了。人们很容易就推测到每个亲代 DNA 分子双链中的任一条链都可作为模板形成子代 DNA 分子的互补链。尽管这个假设一经提出就被认为是最合理可行的，但是仍需有实验的支持。令人高兴的是，在 DNA 双螺旋结构发现后的 5 年内，就有明确的证据显示 DNA 在复制过程中发生了互补链的分离（见第 2 章对 Meselson 和 Stahl 实验的讨论），而且有切实的酶学证据显示单独的 DNA 就可作为新 DNA 链合成的模板。

从某种意义上讲，这些结果使基因如何复制的问题得到了解答；但从另一种意义上讲，对 DNA 复制的研究才刚刚开始。正如我们将在本章中所看到的那样，即使最简单

的 DNA 分子，其复制也是一个复杂的、多步骤的过程。此过程中所涉及的酶，比最初在第一个 DNA 聚合酶发现后所预计的要多很多。真核细胞中大的线性染色体的复制，则更为复杂。这些染色体的复制需要很多复制起始位点，以严格的时相方式合成完整的染色体，而且，复制的起始必须在精确的协调控制下，以确保所有序列都仅仅复制一次。此外，保守性的 DNA 复制不能完全复制染色体末端（端粒），细胞必须形成一套新的方式来保持这部分染色体的完整性。

本章将首先讲述 DNA 合成的化学基础以及催化此反应酶的功能，然后再讨论在一条完整的染色体上 DNA 的合成是如何在复制叉上起始的。在这些位点上的 DNA 复制需要一系列额外的蛋白质。本章最后部分将着重于 DNA 复制的起始和终止。在所有细胞中，DNA 的复制都受到严格的控制，而复制起始是受调控的步骤。我们将描述复制起始蛋白是如何在基因组上被称为复制起始位点的特定位点处解开 DNA 双螺旋的。总之，DNA 复制涉及的蛋白质体系向人们展示了一个复杂的机制，其以令人惊叹的速度、准确性和完整性精确执行着 DNA 复制这一重要的生物学过程。

DNA 合成的化学基础

DNA 合成需要脱氧核苷三磷酸和引物-模板接头

进行 DNA 的合成，必须具备 2 个关键的底物。首先，新的合成需要 4 种脱氧核苷三磷酸——dGTP、dCTP、dATP、dTTP（图 8-1a）。核苷三磷酸通过 2′-脱氧核糖上的 5′羟基连接 3 个磷酰基团。最内侧的磷酰基团（最靠近脱氧核糖的基团）称为 α-磷酸，而中间的和最外侧的基团分别称为 β-和 γ-磷酸。

DNA 合成第二个重要的底物是一具有特定排列的单链 DNA 或双链 DNA（图 8-1b），称为引物-模板接头（primer-template junction）。恰如其名，引物-模板接头有 2 个重要的元件。模板提供指导所有互补脱氧核苷酸添加的单链 DNA。引物是与模板互补但比模板短的一小段序列。引物还必须有一个与模板单链区毗邻的暴露的3′-OH。当新核苷酸加入时此 3′-OH 使链得以延伸。

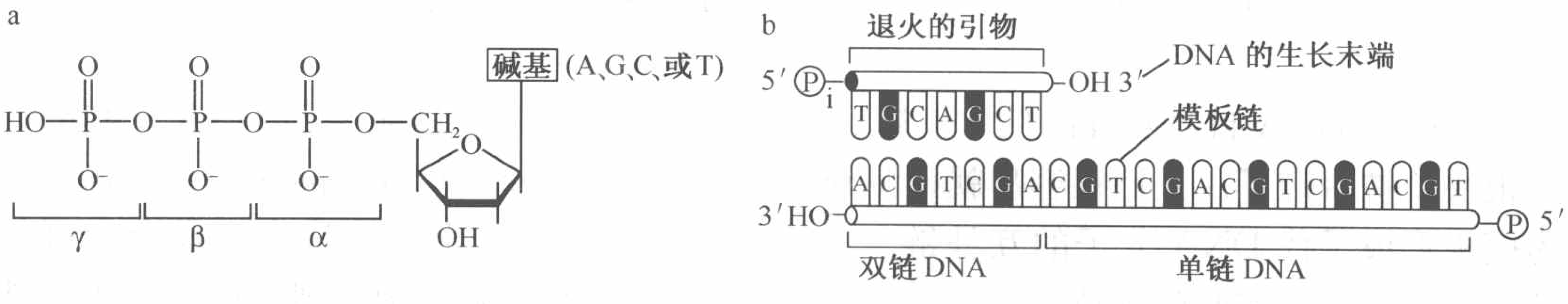


图 8-1 DNA 合成所需的底物。 a. 2′-脱氧核糖核苷三磷酸的常规结构。α-、β-和 γ-磷酸的位置被标示。b. 引物-模板接头的常规结构。较短的引物与较长的 DNA 链完全退火，而且它必须有与模板单链 DNA 区毗邻的游离的 3′-OH；较长的 DNA 链有一个与引物退火的区域和一个毗邻的单链 DNA 区，此区作为 DNA 合成的模板。新 DNA 的合成使引物的 3′端得到延伸。

一般而言，只有引物-模板接头的引物部分才是 DNA 合成的底物，这是因为在 DNA 合成过程中只有引物被化学修饰。模板仅提供选择哪种核苷酸进行添加的必要的信息。然而，对所有 DNA 合成而言，引物和模板都是不可缺少的。

DNA 通过引物 3′端的延伸进行合成

DNA 合成要求新链通过引物 3′端的延伸而生长（图 8-2）。事实上，这是 RNA 和 DNA 合成共同的性质。S_N2 反应中形成磷酸二酯键。在此反应中，引物链 3′端的羟基基团攻击引入的核苷三磷酸的 α-磷酰基基团，离去的基团是焦磷酸。它是核苷酸底物释放 β-和 γ-磷酸所产生的。

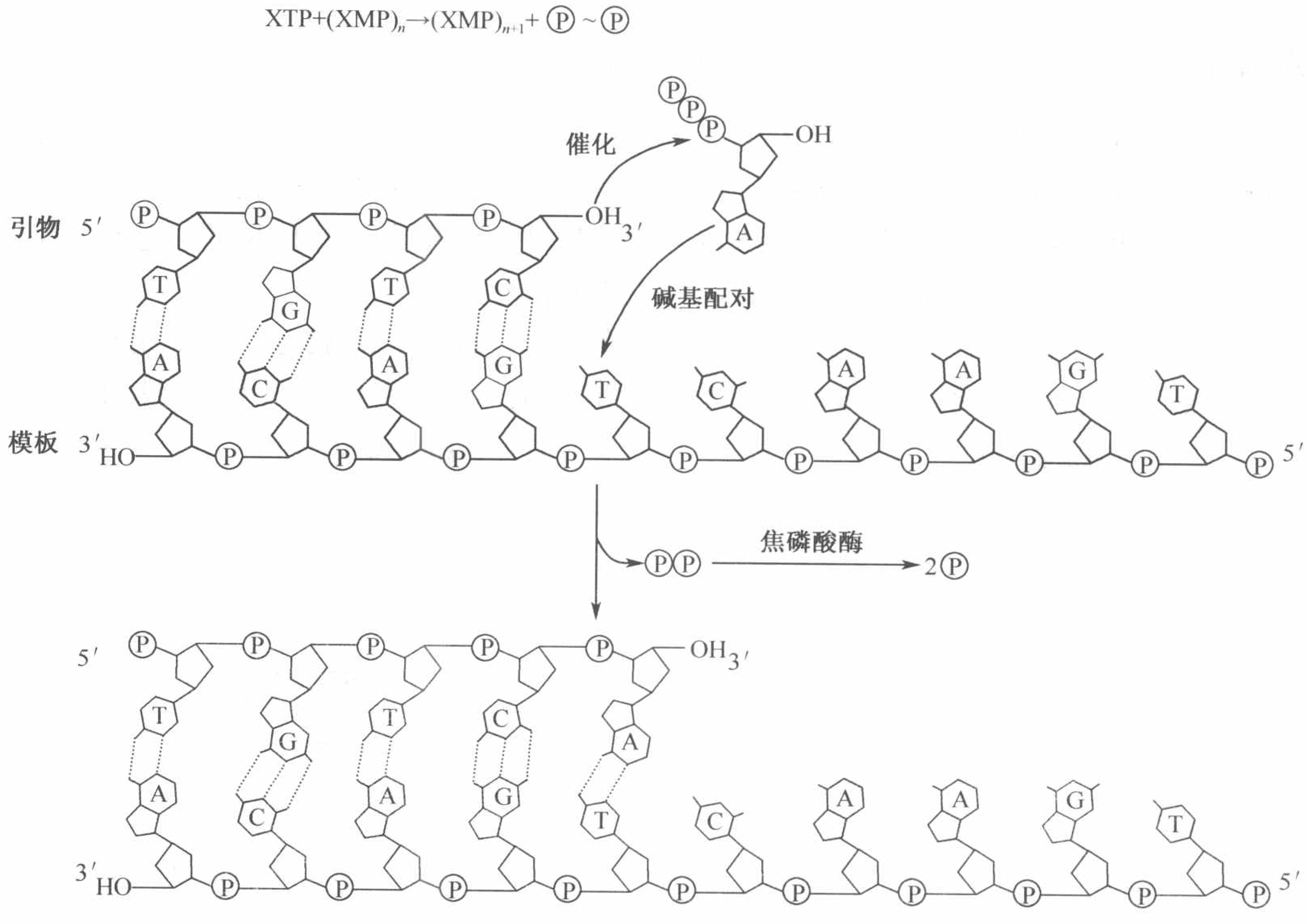


图 8-2 DNA 合成的机制图。 DNA 合成由掺入 dNTP 的 α-磷酸的亲核攻击引发。这导致引入的引物 3′端延伸一个核苷酸并释放一个焦磷酸分子。焦磷酸很快被焦磷酸酶水解成 2 个磷酸分子。

模板链指导 4 种核苷三磷酸的添加。与模板链碱基配对的核苷三磷酸非常容易被添加到引物链上。注意双螺旋的两条链为反向平行的走向。这种结构意味着 DNA 合成的模板链有与 DNA 生长链相反的走向。

焦磷酸水解是 DNA 合成的驱动力

单个核苷酸添加到长度为 n 的多聚核苷酸生长链上可用下面的反应来表述：

$$\text{XTP} + (\text{XMP})_n \rightarrow (\text{XMP})_{n+1} + \text{Ⓟ} \sim \text{Ⓟ}$$

但是此反应的自由能非常小（$\Delta G = -3.5$ kcal/mol），那么核苷酸多聚化形成 DNA 的

驱动力是什么呢？额外的自由能由焦磷酸酯酶将焦磷酸快速水解成 2 个磷酸基团来提供。

$$P \sim P \rightarrow 2P_i$$

核苷酸添加和焦磷酸水解的净结果是 2 个高能磷酸键的断裂。因此，DNA 合成是个耦联反应，其全反应为

$$XTP + (XMP)_n \rightarrow (XMP)_{n+1} + 2P_i$$

这是一个极易发的反应，其 ΔG 为 -7kcal/mol，相对应的平衡常数（K_{eq}）约为 10^5。如此高的 K_{eq} 意味着 DNA 合成反应是完全不可逆的。

DNA 聚合酶的作用机制

DNA 聚合酶用一个活性位点催化 DNA 的合成

DNA 合成由 DNA 聚合酶（DNA polymerase）催化。与大多数对一个反应就有一个活性位点的酶不同，DNA 聚合酶用一个活性位点来催化任意 4 种脱氧核苷三磷酸的添加。DNA 聚合酶通过利用 A：T 和 G：C 碱基对几乎相当的几何尺寸来达到其催化的灵活性（注意 DNA 螺旋的尺寸与 DNA 序列基本无关）。

DNA 聚合酶监视引入核苷酸形成 A：T 或 G：C 碱基对的能力，而不是检测进入活性位点的是何种核苷酸（图 8-3）。只有在形成正确的碱基对的情况下，引物的 3′-OH

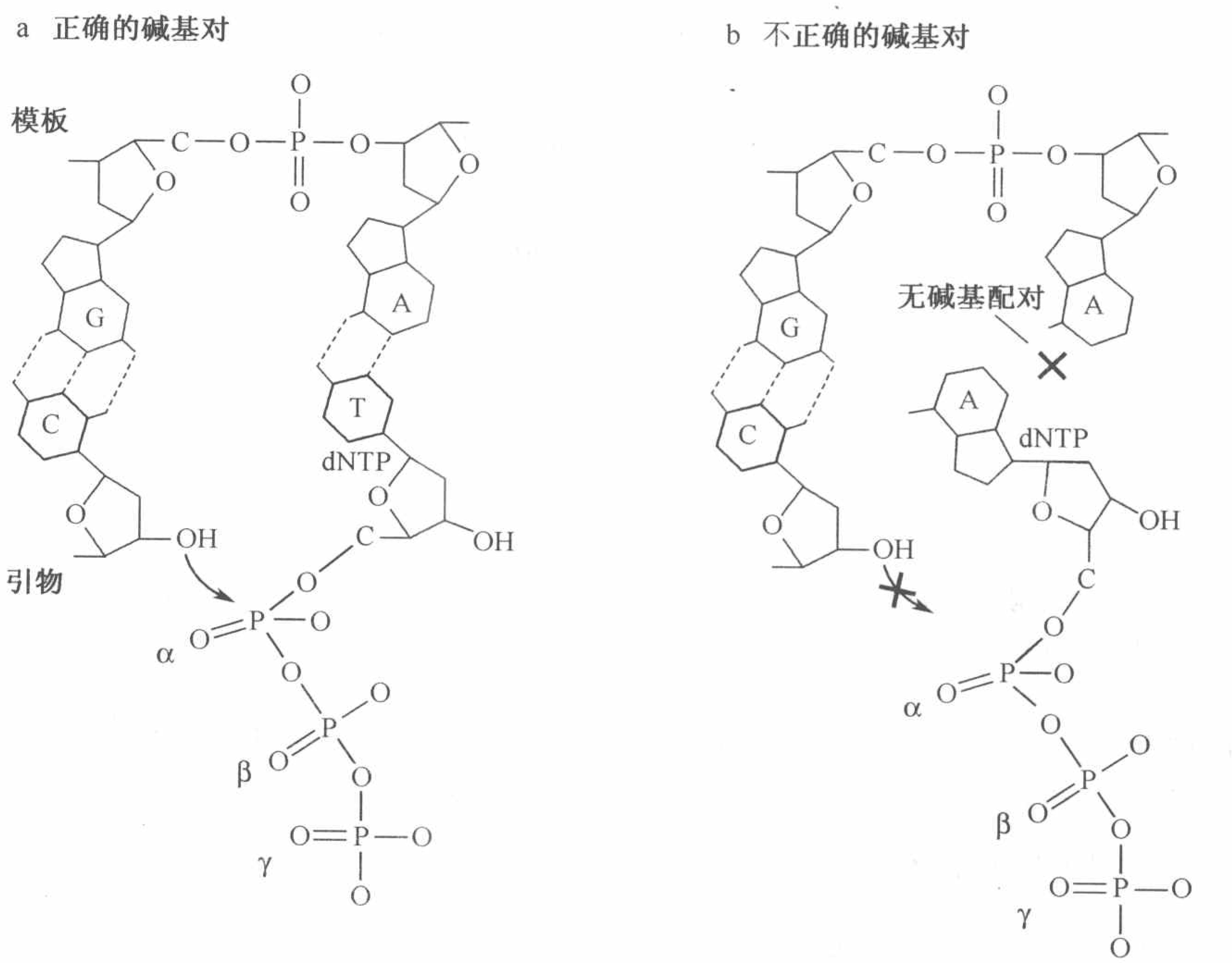


图 8-3 DAN 聚合酶催化核苷酸添加需要正确配对的碱基。 a. 引物 3′-OH 端对正确配对的 dNTP 进攻的示意图。b. DNA 聚合酶对不正确碱基配对催化结果的示意图。如图所示，不正确的 A：A 碱基对使引入核苷酸的 α-磷酸错位。这种不正确的排列极大地降低了催化速率，导致 DNA 聚合酶优先添加正确配对的 dNTP。（来源：Based on Brautigan C. A. and Steitz T. A. 1998. Structural and functional insights provided by crystal structures of DNA polymerase. Curr. Opin. Structural Biology 8：60，Fig 4，d. © Elsevier.）

和在最佳催化位置上的引入核苷三磷酸的 α-磷酸才能发生催化反应。不正确的碱基配对因为使底物处于不利于催化的排列，使得核苷酸的添加效率显著降低（图 8-3b）。这是一个动力学选择性的实例。其中，酶对几种可能的底物具有催化选择性，只有当正确的底物存在时，共价键形成的速率才显著增加。事实上，错误核苷酸掺入的速率是碱基配对正确时的掺入速率的 1/10 000。通常监视新 DNA 合成的方式描述如框 8-1，核素掺入法能够被用来测定核苷酸和蛋白质合成。

DNA 聚合酶对核糖核苷三磷酸（rNTP）与脱氧核糖核苷三磷酸（dNTP）显示出令人惊异的分辨能力。虽然在细胞中 rNTP 比 dNTP 的浓度要高约 10 倍，但它们掺入的速率却是 dNTP 的 1/1000。这种分辨能力是通过将 rNTP 立体排斥在 DNA 聚合酶活性位点之外而达到的（图 8-4）。在 DNA 聚合酶中，核苷酸结合口袋非常小，不能容纳引入核苷酸上的 2′-OH。此空间被 2 个与糖环形成范德华接触的氨基酸占据。有趣的是，将这些氨基酸换成其他较小侧链的氨基酸（如将谷氨酸换成丙氨酸），则 DNA 聚合酶对 dNTP 和 rNTP 的分辨率显著降低。符合 DNA 聚合酶一些要求但不是所有要求的核苷酸能够阻止 DNA 聚合反应的延伸。这些核苷酸也代表了一类用于治疗癌症和病毒感染的药物。（见框 8-2，抗癌，抗病毒药物靶向 DNA 复制）

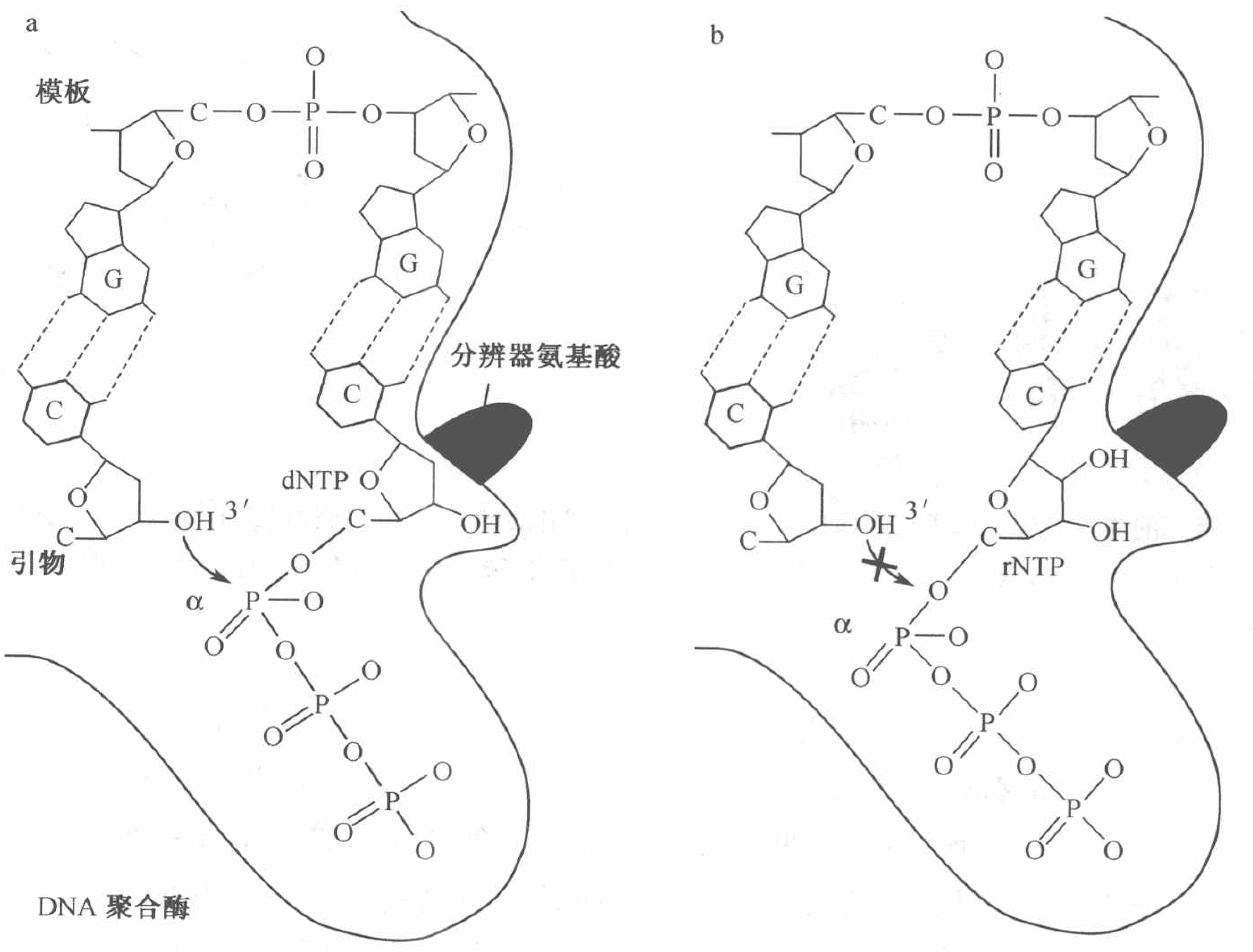
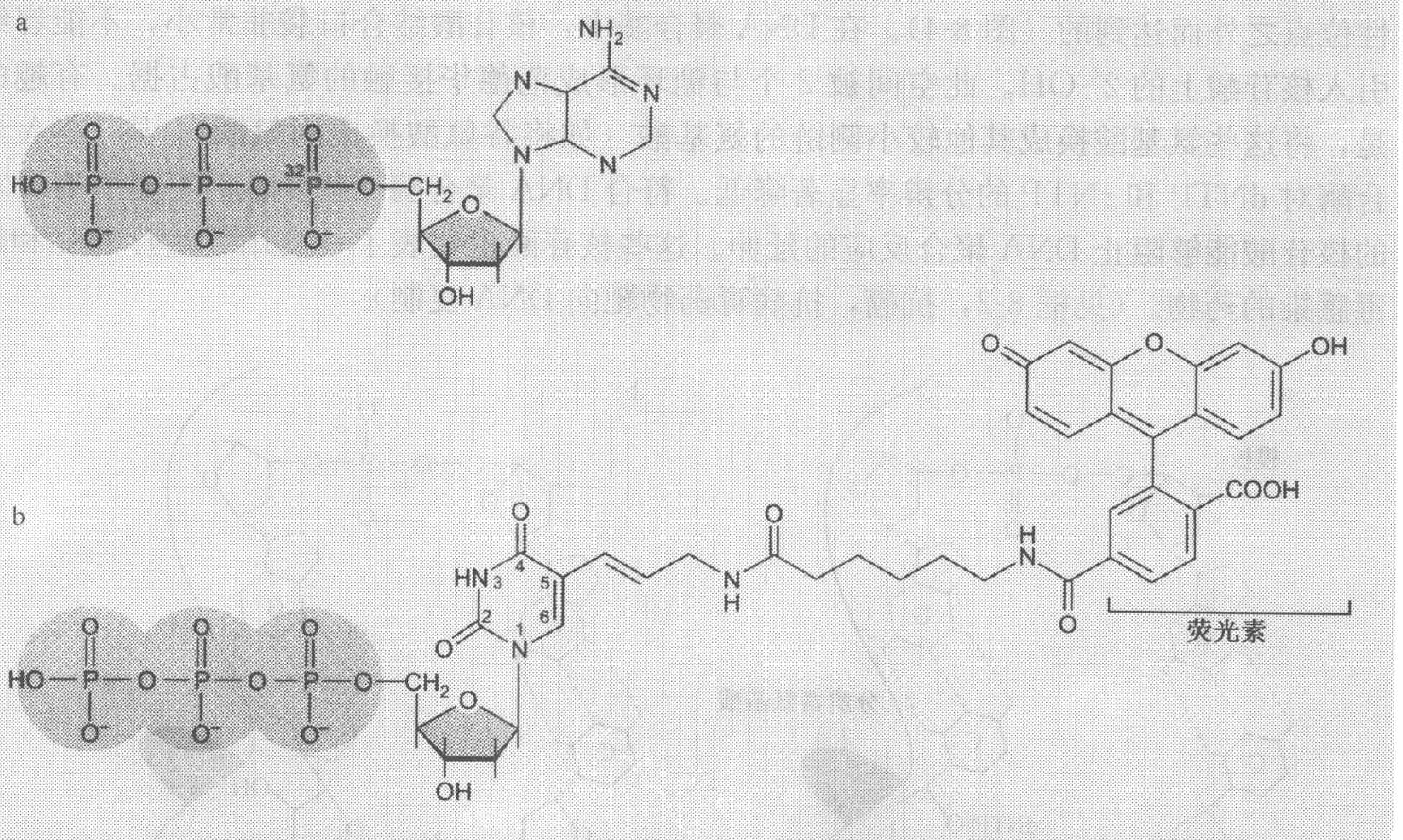


图 8-4　空间位阻阻止 DNA 聚合酶催化反应使用 rNTP。 a. 正确碱基配对的 dNTP 结合到 DNA 聚合酶上，在此条件下，引物的 3′-OH 与 dNTP 的 α-磷酸紧密靠近。b. 增加 2′-OH 引起对核苷酸结合口袋内氨基酸［分辨器（discriminator）氨基酸］的空间碰撞。导致 dNTP 的 α-磷酸错位，以及与引物 3′-OH 的定位偏差（misalignment），极大地降低了催化的速率。

框 8-1 核素掺入法用来测定核酸和蛋白质合成

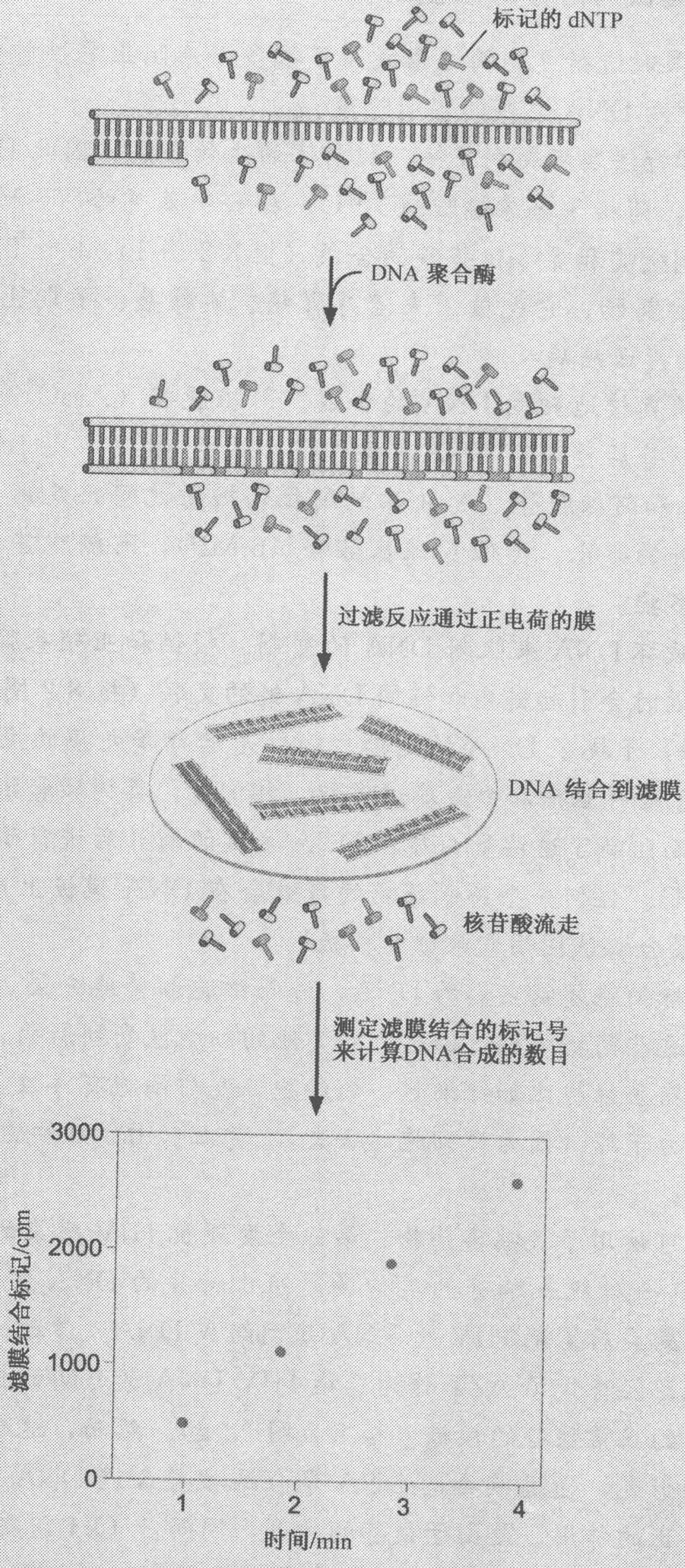
DNA 聚合酶的活性如何测定？最简单的用于测定一种聚合物合成的方法是核素掺入法。在 DNA 聚合酶实例中，用这种方式来测定被标记的 dNTP 前体整合进 DNA 分子。典型的实例，dNTPs 被标记成含有放射性原子，这些放射性原子最终保留在 DNA 产物中（例如用放射性同位素^{32}P 替换 α 位磷酸中的磷原子，框 8-1 图 la)。或者，dT-TP 甲基基团位置的核苷酸用荧光分子合成（框 8-1 图 lb)。甲基基团不包含在碱基配对中，DNA 聚合酶能够容易地适应这个位点比较大的成分。另外一种情况，这些修饰能够被容易地用拍照或敏感的光电倍增管探测电子或光子的情况。



框 8-1 图 1 标记的脱氧核苷三磷酸盐。(a) α［^{32}P］dATP. 核苷酸的 α 位的 P 被替代成了放射性同位素^{32}P。注意仅有 P 原子最终会成为合成后 DNA 的一部分。(b) 荧光标记的三磷酸胸苷。在这个标记的前体中，荧光复合物荧光素和胸腺嘧啶环的第五位相连，这个位置通常连接的是甲基。

核素掺入法需要两步（框 8-1 图 2)。第一步，前体被整合进聚合物。在 DNA 聚合酶的实例中，这些通过用引物产生聚合物来实现：模板结合和适当时期标记的 dNTP 前体。在大多数情况下，只有四种 dNTPs 中的一种被标记。第二步，产生的聚合物必须和没有整合进的前体分离开。以 DNA 为例，这一步有两种方法完成。一种方法是，DNA 聚合酶反应能够在盐浓缩环境中通过带正电荷过滤器，带强负电荷的 DNA 主链能够结合上去，而不是弱电荷的单个核苷酸。再者，凝胶电泳能够用来通过大小区分开 DNA 产物。另外一种情况，DNA 产物的合成能够用标记的核苷酸整合进 DNA 聚合物中的数目来测定。

我们描述了以 DNA 聚合反应为背景的核酸掺入法；但是，这种方法也能够用来测定催化 RNA 和蛋白质的合成的酶的活性。比如，标记的氨基酸能够用来分析它们整合进蛋白质的情况。



框 8-1 图 2　核素掺入法来测定 DNA 合成。如例所示，滤膜结合用来分离没有被结合的核苷酸和合成的 DNA 链。

框 8-2 抗癌和抗病毒试剂靶向 DNA 复制

DNA 在细胞分裂时进行复制的过程成为了很多用来阻止恶性肿瘤生长的化学药物的靶向。这些药物靶向 DNA 复制的不同阶段。

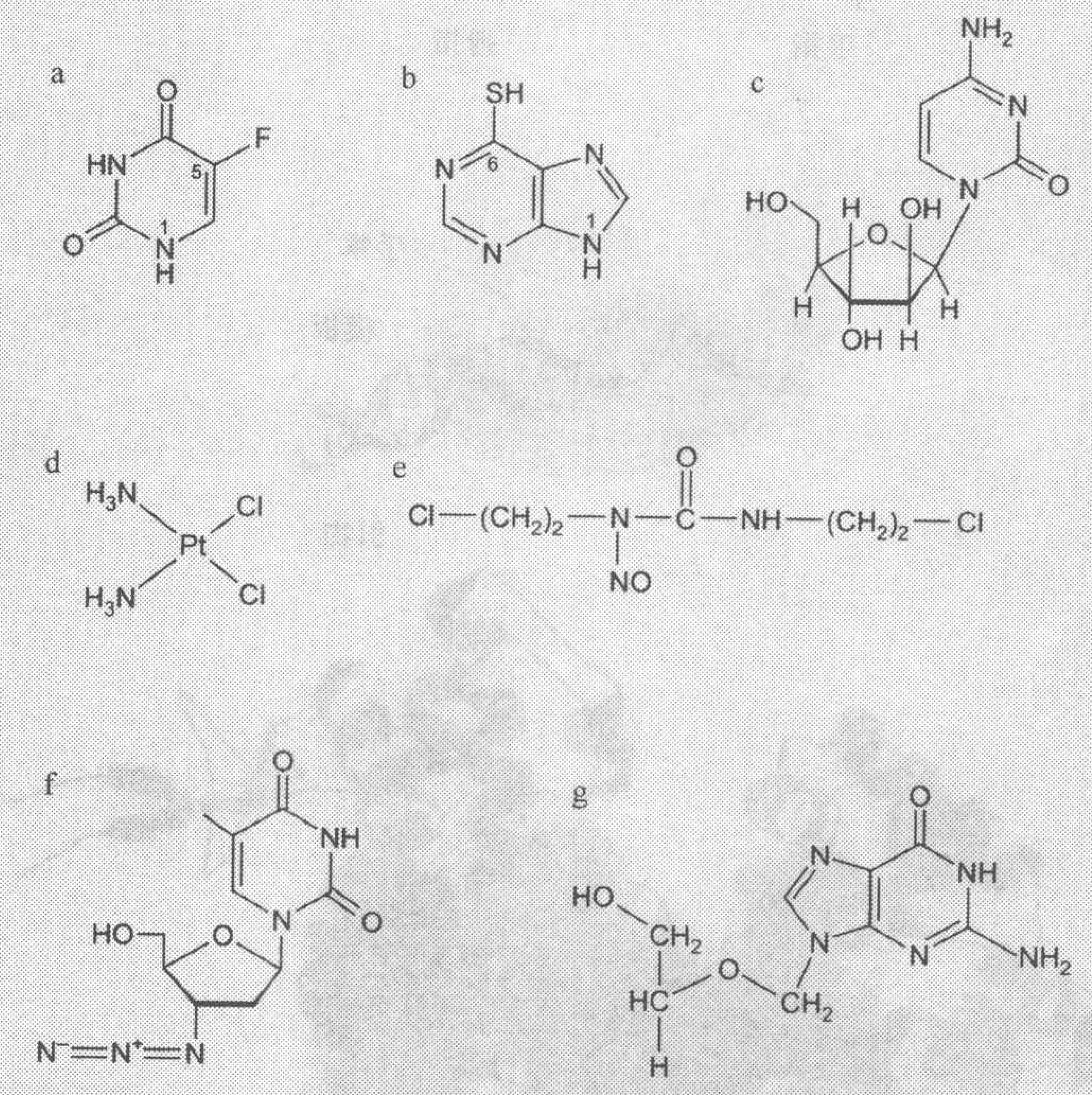
一些常用的化学疗法实际靶向了核苷酸前体的生物合成，因此 DNA 聚合酶没有底物合成 DNA。比如，药物 5-氟尿嘧啶（5-FU）和 6-巯基嘌呤（6-MP）是核苷酸前体的类似物，分别抑制嘧啶和嘌呤核苷酸的合成（框 8-2 图 1a，b）。5-氟尿嘧啶（5-FU）是治疗结肠癌的土要药物，它也被用来治疗胃癌、胰腺癌、早期乳腺癌。6-巯基嘌呤（6-MP）基本用于治疗血癌病人。

其他的抗癌药物直接地靶向 DNA 的合成。阿糖胞苷（AraC）是脱氧胞苷酸类似物，转化成三磷酸核苷后整合进 DNA，替代了 dCTP（框 8-2 图 1c）。一旦整合进去，dCTP 脱氧核糖糖基和阿糖胞苷（AraC）的阿拉伯糖基之间的差异导致 DNA3′端位置的不正确和末端延伸的终止。类似 6-巯基嘌呤（6-MP），阿糖胞苷（AraC）主要用于治疗急性白血球过多症。

第三类的药物破坏 DNA 来阻断 DNA 的复制。顺铂和亚硝基脲氮芥（BCNU）在鸟苷酸残基彼此相近时会引起链内和链间 DNA 链的交联（框 8-2 图 1d，e）。这种交联（尤其是链内的交联）干扰了 DNA 的延伸。顺铂是治疗睾丸癌的主要药物，亚硝基脲氮芥（BCNU）用于治疗脑瘤和白血球过多症。相似地，喜树碱和依托泊苷对拓扑异构酶有抑制作用，进而阻断了这些蛋白质在 DNA 主链的断开后进行磷酸二酯结合的作用（见第 6 章，图 6-24）。任何一个抑制药物的疗效会在 DNA 模板上产生阻断 DNA 复制的断裂，而 DNA 聚合酶试图用它来作为模板。

这些药物的靶标细胞复制它们的 DNA，进而细胞频繁地分裂。尽管癌细胞快速分裂的特性使它们对这些药物敏感，但是体内其他的细胞也受到影响。毫无疑问，这些药物也对快速分裂的宿主细胞比如红细胞、白细胞、发质细胞有毒性。使用这些药物带米的副作用包括免疫力下降（因为白细胞的丢失）、贫血（由于红细胞的丢失）、腹泻（肠胃的缺陷）和脱发。

复制抑制剂也可被用于抗病毒药物。第一个发现抗 HIV 感染的有效药物是 3-叠氮脱氧胸苷（AZT），一种脱氧胸苷的类似物，阻止特异的 DNA 聚合酶（称逆转录酶；见 11 章）在 HIV 感染后复制 HIV 的 RNA 基因组成 DNA。最近一种鸟嘌呤核苷酸类似物叫做阿昔洛韦已经替代了 AZT 作为感染 HIV DNA 聚合酶的抑制剂。这个类似物是具有一个开放链的正常腺苷的核糖（框 8-2 图 f，g）。然而，这个类似物能够被修饰成一个三磷酸根的形式，通过病毒的 DNA 聚合酶被整合进 DNA。一旦整合进去，这些类似物起到链终止的作用，是由于核苷的缺乏，因而 3′-OH 需要更多的核苷酸添加。重要的一点是，这些药物被细胞内的 DNA 聚合酶很弱地识别，因而同化学疗法试剂相比具有很小的副作用。



框 8-2 图 1　靶向 DNA 复制的一些常见化学疗法试剂的结构。(a) 5-氟尿嘧啶；(b) b-巯基嘌呤；(c) 阿糖胞苷；(d) 顺铂；(e) 亚硝基脲氮芥；(f) 叠氮胸苷；(g) 阿昔洛韦

DNA 聚合酶像手一样握住引物：模板接头

关于 DNA 聚合酶如何催化 DNA 合成的分子解释，来自于对各种与引物-模板接头结合的 DNA 聚合酶的原子结构的研究。这些结构揭示了 DNA 底物位于一个大的、形似一只半握的右手的裂缝之中（图 8-5）。依据对手的类比，聚合酶的 3 个结构域被称为拇指、手指和手掌。

手掌域由一个 β 折叠片构成，含有催化位点的基本元件，尤其是 DNA 聚合酶的这一区域结合 2 个二价金属离子（通常是 Mg^{2+} 或 Zn^{2+}），可以改变正确碱基配对的 dNTP 和引物 3′-OH 周围的化学环境（图 8-6）。一个金属离子降低 3′-OH 对其氢的亲和力，这会产生一个准备对引入 dNTP 上 α-磷酸进行亲核攻击的 $3'O^-$。第二个金属离子与 dNTP 的 β-和 γ-磷酸负电荷协同作用，稳定由引物和引入核苷酸连接在一起时所产生的焦磷酸。

除了在催化中的作用外，手掌域还负责检查最新加入的核苷酸碱基配对的准确性。聚合酶的这一区域与新合成 DNA 小沟中的碱基对形成大量的氢键。这种接触虽无碱基特异性，但是只有在新添加的核苷酸（不管它们是哪一种核苷酸）碱基配对正确的时候才形成。此区域中错误配对的 DNA 使催化活性显著降低。变慢的催化反应和对新合成 DNA 降低的亲和力共同使引物-模板从聚合酶活性位点上脱离下来，并结合到聚合酶上的一个独立的具校对功能的核酸酶的活性位点上。

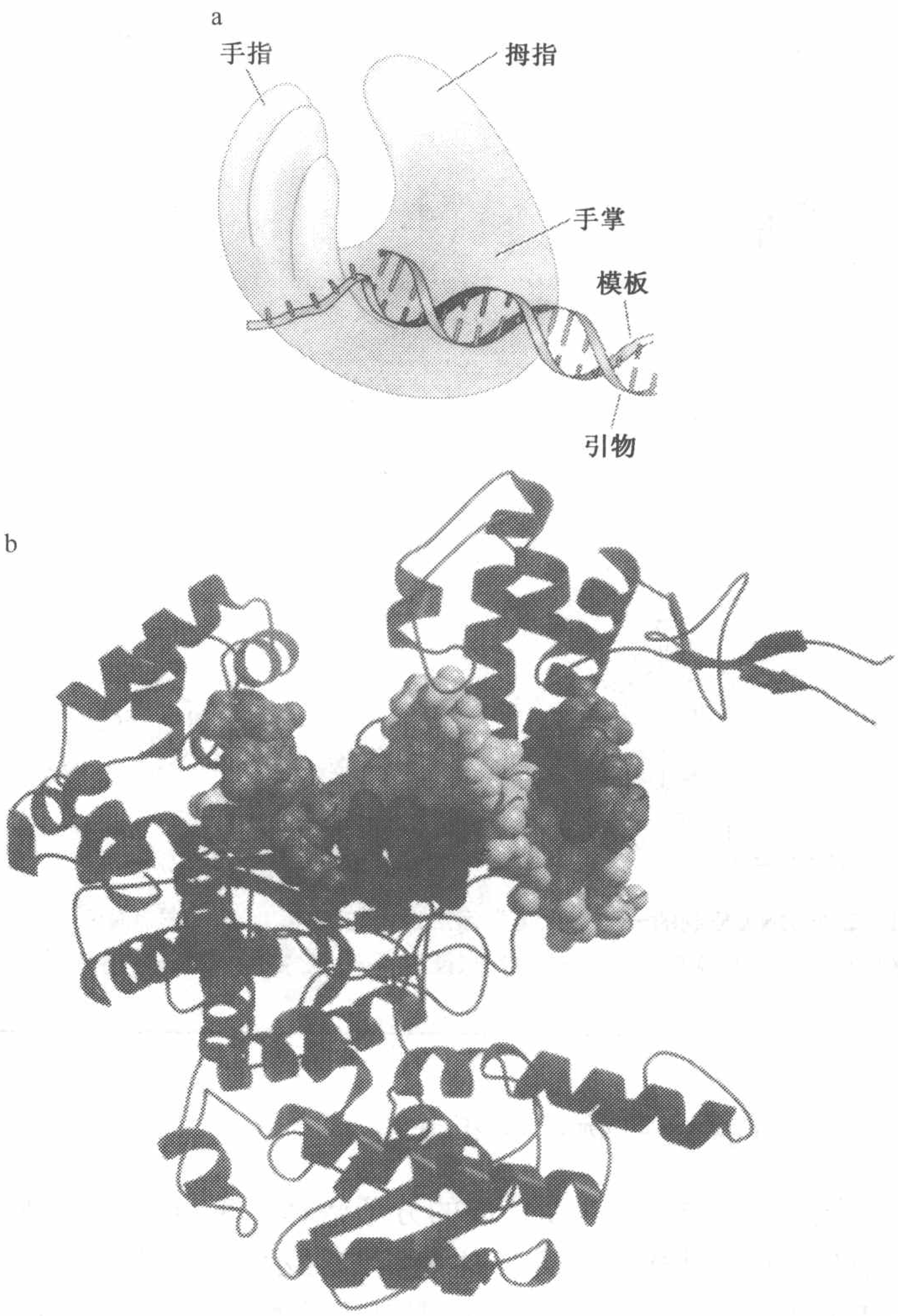


图 8-5 DNA 聚合酶的三维结构类似于一只右手。 a. 图示 DNA 聚合酶与引物-模板接头结合。手指、拇指和手掌已被标出。新合成的 DNA 与手掌相连，DNA 催化位点位于手指和拇指之间的裂缝中。模板链的单链区强烈弯曲，不从拇指和手指之间穿过。b. 类似的 T7 DNA 聚合酶与 DNA 结合的图示。DNA 用填充（space-filling）方式表示，蛋白质用条带（ribbon）方式显示。手指和拇指由 α 螺旋构成，手掌因为 DNA 的遮挡而看不见。引入的 dNTP 显示为红色（碱基和脱氧核糖）和黄色（三磷酸部分）。DNA 模板链显示为黑灰色，引物链显示为银灰色。（Doublie S.，Tabor S.，Long A. M.，Richardson C. C.，and Ellenberger T. 1998. Nature 391：251.）镜像图用 BobScript、MolScript 和 Raster 3D 制作。

手指域和拇指域的作用是什么？手指域对催化也很重要。手指域中的几个残基可与引入的 dNTP 结合。更重要的是，一旦引入的 dNTP 与模板之间形成正确的碱基配对，手指域即发生移动，包围住 dNTP（图 8-7）。聚合酶手的这种闭合形式可以使引入的核苷酸与催化性的金属离子密切接触，从而促进催化反应。

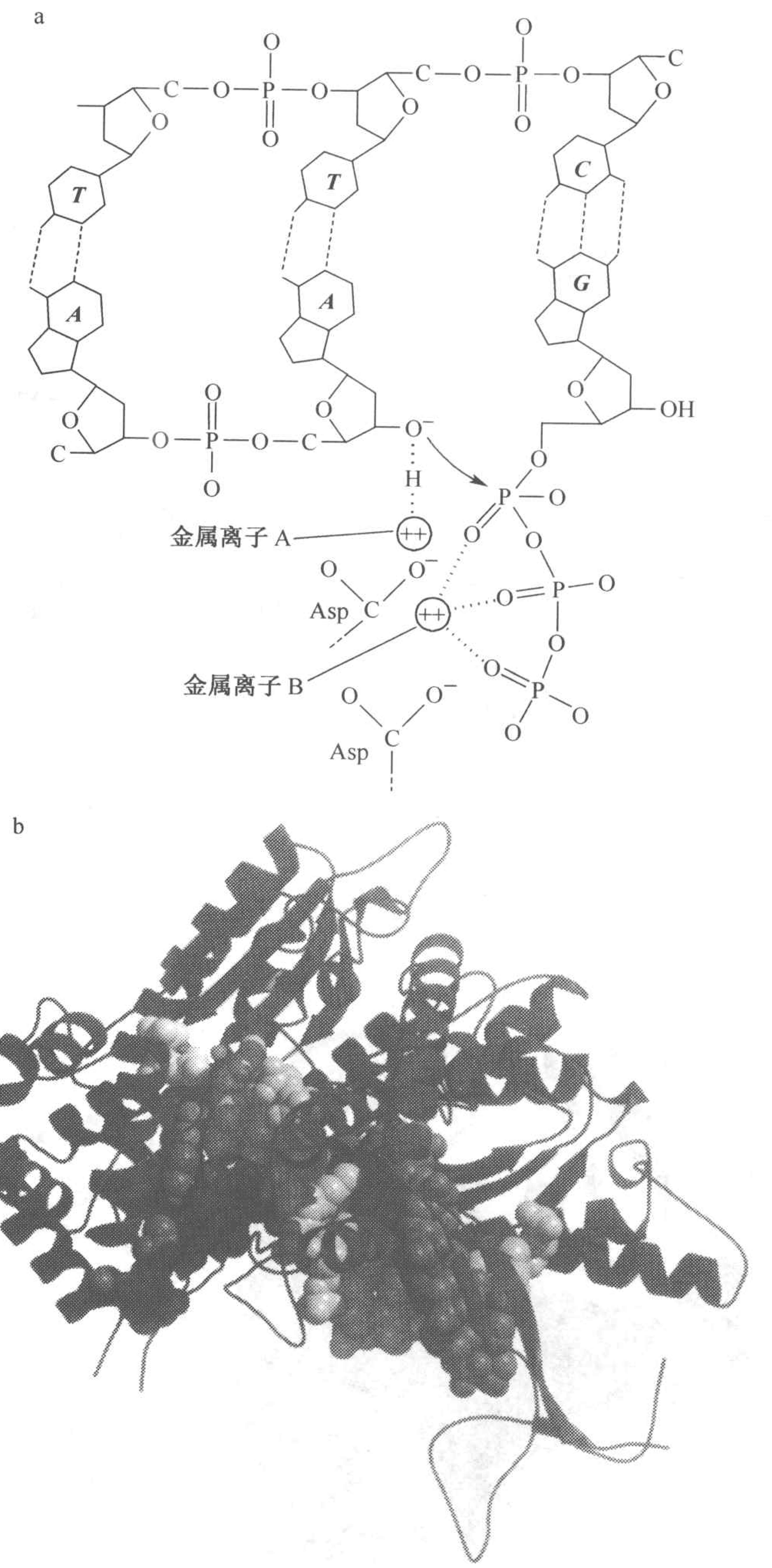
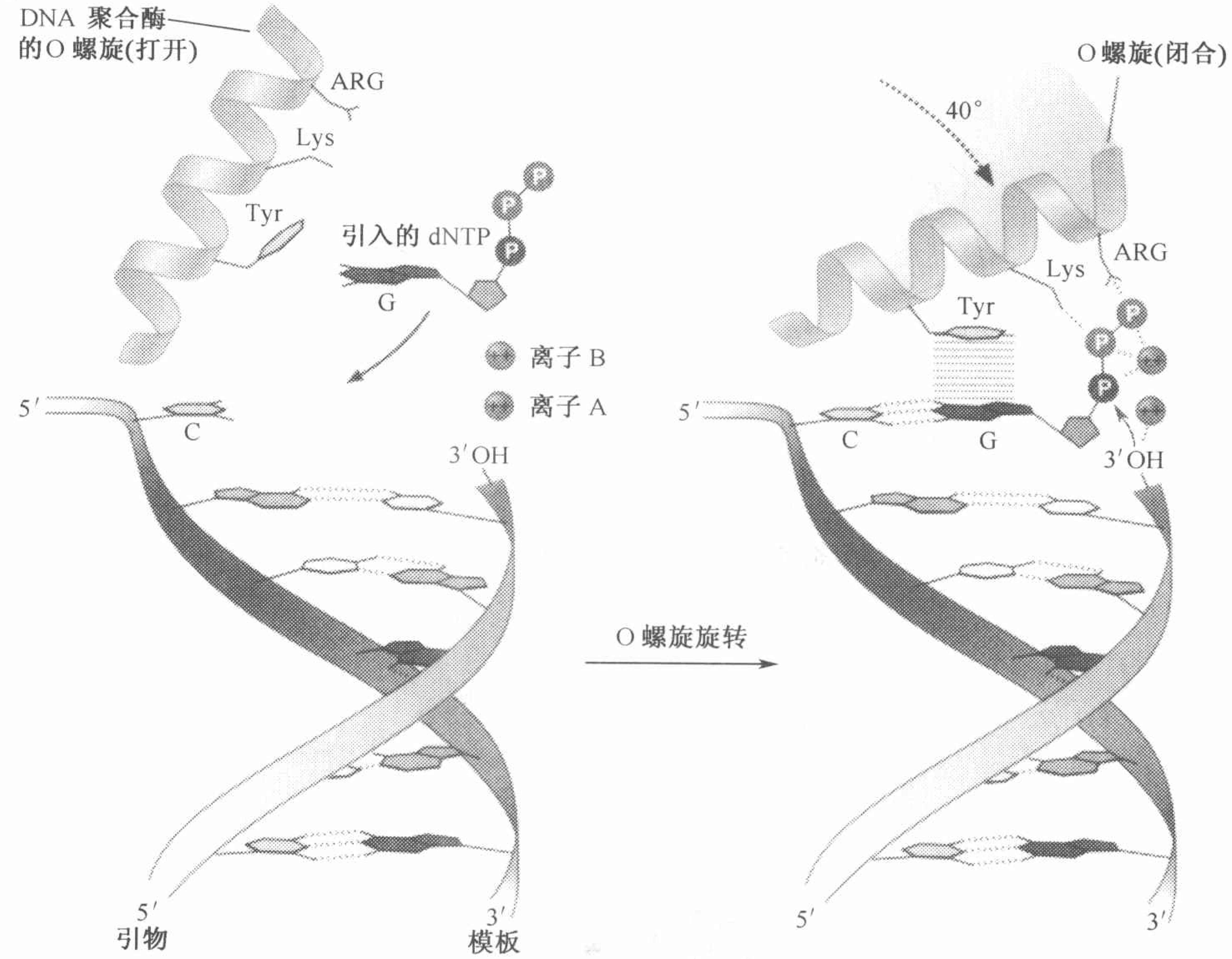


图 8-6　与 DNA 聚合酶结合的两个金属离子催化核苷酸的添加。 a. 图示为 DNA 聚合酶的活性位点。两个金属离子（显示为绿色）通过与两个高度保守的天冬氨酸残基的相互作用被固定。金属离子 A 主要与 3′-OH 相作用，使得 O 和 H 之间的连接变弱。这产生一个亲核的 3′O⁻；金属离子 B 与引入的 dNTP 的三磷酸相互作用，中和它的负电荷。催化后，焦磷酸产物通过与金属离子 B 类似的相互作用得到稳定（未显示）。b. 金属离子与 DNA 聚合酶、引物 3′-OH 端以及引入核苷酸的相互作用的活性位点处的三维结构。金属离子显示为绿色，其他元件的颜色同图 8-5b。本处所示的聚合酶的视角，基本等同于将图 8-5b 中的图沿 DNA 螺旋的轴旋转约 180°的图像。(Doublie S.，Tabor S.，Long A. M.，Richardson C. C.，and Ellenberger T. 1998. Nature 391:251.)镜像图用 BobScript、MolScript 和 Raster 3D 制作。

a



b

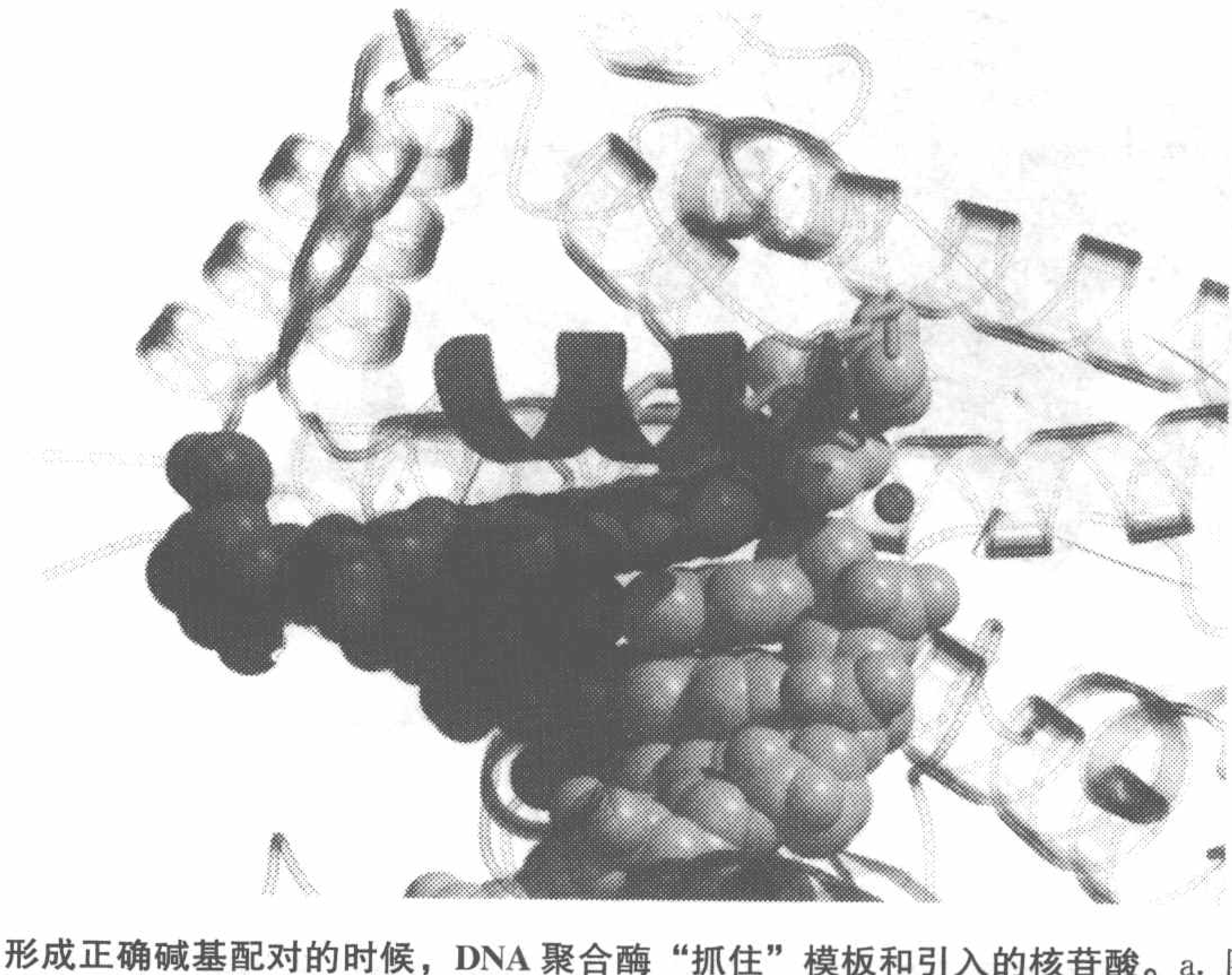


图 8-7 当形成正确碱基配对的时候，DNA 聚合酶“抓住”模板和引入的核苷酸。 a. 图示为引入核苷酸与 DNA 模板正确碱基配对之后，DNA 聚合酶结构的变化。主要的变化是在手指域中，一个被称为 O 螺旋的螺旋旋转 40°。在打开的构象中，此螺旋远离引入的核苷酸。当聚合酶为闭合构象时，此螺旋运动并与引入的 dNTP 形成几个重要的相互作用。酪氨酸与 dNTP 的碱基形成堆积作用，两个带电残基与三磷酸结合。这些相互作用共同使 dNTP 就位，并由与 DNA 聚合酶结合的两个金属离子介导完成催化。b. 与底物结合的 T7 DNA 聚合酶处于闭合构象时的结构。O 螺旋显示为紫色；蛋白质结构中的其他部分为清楚起见显示为透明；在 O 螺旋后面的关键的酪氨酸、赖氨酸和精氨酸显示为粉色；引入 dNTP 的碱基和脱氧核糖显示为红色；引物显示为银灰色；模板链显示为黑灰色；两个催化性金属离子显示为绿色；磷酸显示为黄色。(Doublie S.，Tabor S.，Long A. M.，Richardson C. C.，and Ellenberger T. 1998. Nature 391:251.)镜像图用 BobScript、MolScript 和 Raster 3D 制作。

手指域还与模板区结合，使模板的磷酸二酯键骨架在活性部位后立即产生约 90°的回折，此弯曲仅使催化位点上引物后第一个模板碱基暴露。模板的这种构象避免了在下一个核苷酸添加时可能会造成的对配对的模板碱基选择的混淆（图 8-8）。

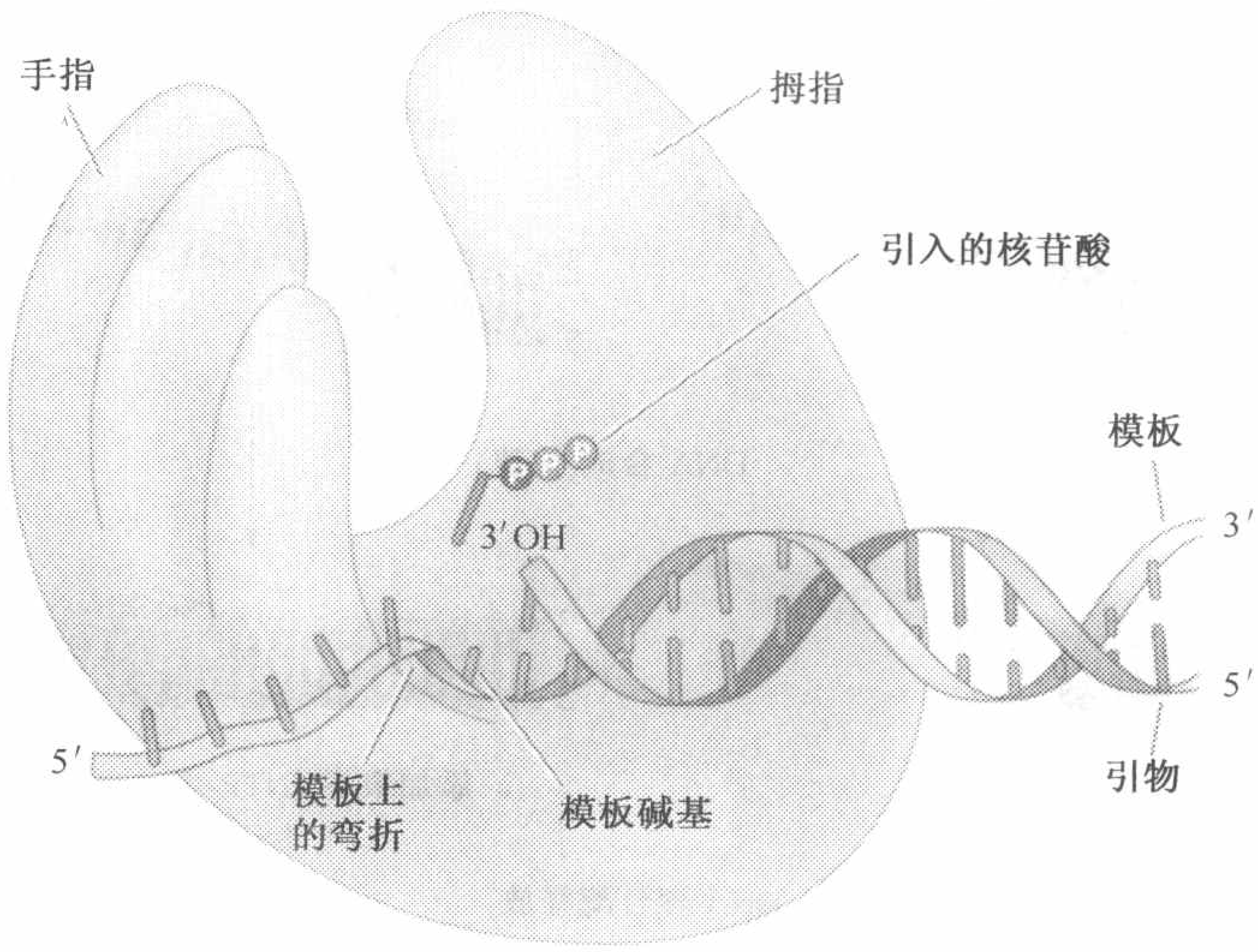


图 8-8 模板 DNA 穿过 DNA 聚合酶的路径示意图。刚合成的 DNA 与 DNA 聚合酶的手掌域结合。在活性位点上，模板单链区的第一个碱基位于预计将合成为双链 DNA 的位置上。当沿着模板链向其 5′端行进时，磷酸二酯骨架突然 90°回折。使得第二个以及随后所有的单链碱基都处于一个位置上，此位置消除了与活性位点上结合的 dNTP 进行碱基配对的所有可能。

与手指域及手掌域不同的是，拇指域与催化的关联不紧密，而是与最新合成的 DNA 相互作用（图 8-9）。这有两个目的：第一，维持引物以及活性部位的正确位置；第二，帮助维持 DNA 聚合酶与其底物之间的紧密连接。这种连接有助于 DNA 聚合酶每次与引物-模板接头结合时，具有添加许多 dNTP 的能力。

总之，DNA 聚合酶将核苷酸添加到 DNA 生长链是一系列有序的事件。引入的核苷酸碱基与模板提供的下一个碱基进行配对，这种作用使聚合酶的“手指”在碱基配对的 dNTP 周围闭合。聚合酶的这种构象将关键的催化性金属离子置于催化下一个磷酸二酯键形成的位置上。碱基配对的核苷酸与引物的结合使手指重新打开，并使引物-模板接头移动一个碱基，然后聚合酶准备进行下一次的添加。重要的是，这些事件的每一步都由引入 dNTP 与模板之间正确的碱基配对强烈推动。

DNA 聚合酶是一种延伸酶

DNA 聚合酶的催化是快速的。DNA 聚合酶能在引物链上每秒添加 1000 个核苷酸。DNA 合成的速度主要取决于 DNA 聚合酶的延伸能力。延伸能力（processivity）是酶处理多聚体底物时的一种特性。对于 DNA 聚合酶，其延伸能力定义为每次酶与引物-模板接头结合时所添加核苷酸的平均数。每个 DNA 聚合酶都有其特征性的延伸能力，范围从每次结合时仅几个核苷酸到 5000 多碱基（图 8-9）。

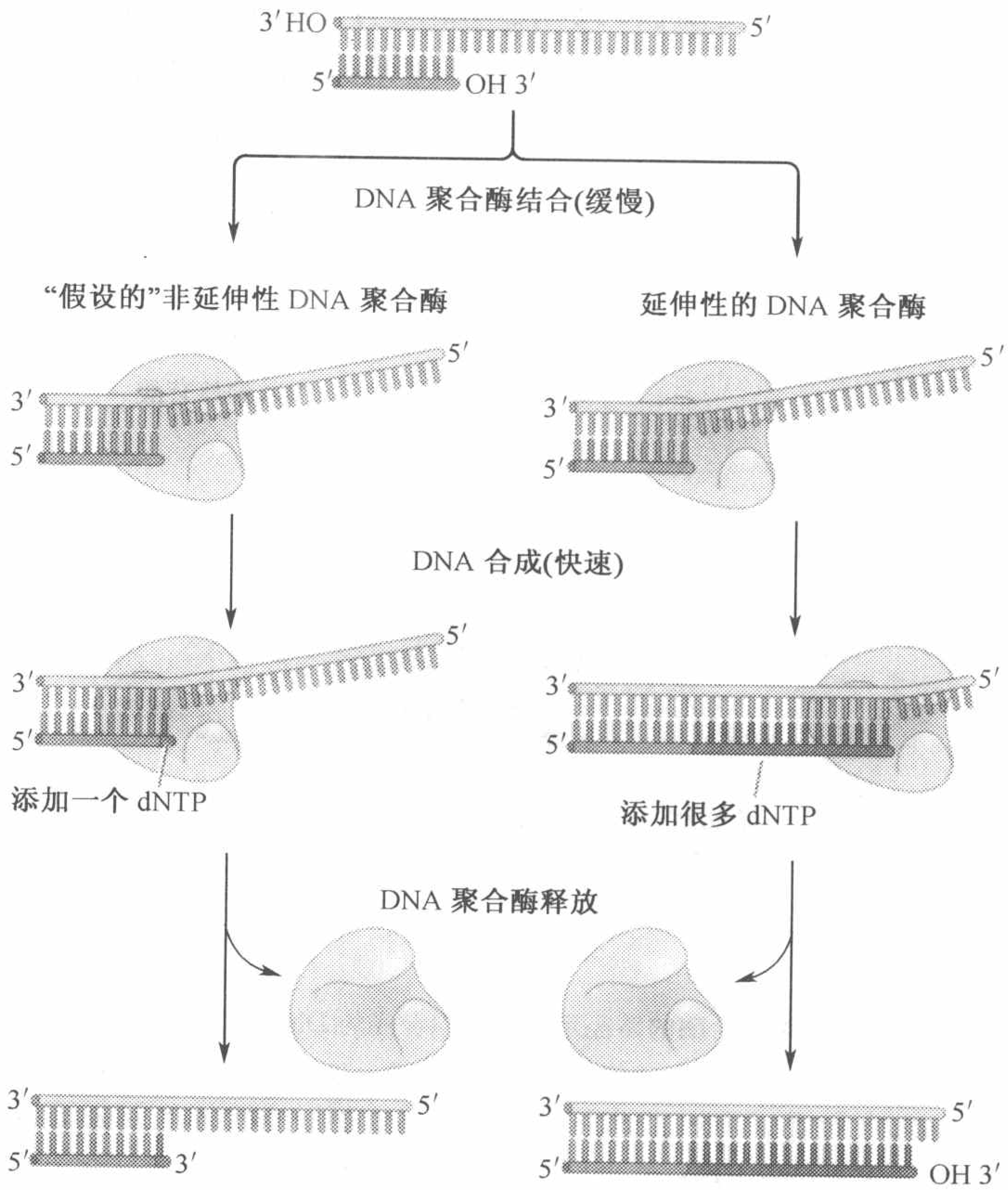


图 8-9 DNA 聚合酶以延伸方式合成 DNA。 此图表示延伸性 DNA 聚合酶和非延伸性 DNA 聚合酶之间的差别。两种 DNA 聚合酶都与引物-模板接头结合。当结合时，非延伸性酶在引物的 3′端添加一个 dNTP，然后从新的引物-模板接头释放下来。相反，延伸性的 DNA 聚合酶与模板结合一次，就添加很多 dNTP。

每次结合添加大量的核苷酸，使 DNA 合成的速率显著提高。聚合酶与引物-模板接头的最初结合是限速步骤。典型的 DNA 聚合酶反应中，DNA 聚合酶定位并结合到引物-模板接头的时间约为 1s，一旦结合后，核苷酸的添加就非常迅速了（在毫秒范围内）。因此，一个完全非延伸性的 DNA 聚合酶每秒添加约 1 个碱基对；相反，最快的 DNA 聚合酶通过保持与模板的结合进行多轮 dNTP 的添加，速度可达每秒添加 1000 个核苷酸。因此，高延伸性的聚合酶与一个完全非延伸性的酶相比，其 DNA 合成的整体速率可高出 1000 倍。

DNA 聚合酶沿 DNA 模板滑行的能力有利于提高延伸能力。DNA 聚合酶一旦与引物-模板接头结合，立即以序列非特异的方式与 DNA 的大部分双链区域紧密作用。此相互作用包括磷酸骨架与拇指域之间的电子静电相互作用，以及 DNA 小沟与手掌域（如上所述）之间的相互作用。此相互作用无序列相关性，使得 DNA 即使在与聚合酶相结合后也能自如地移动。每次引物链上添加一个核苷酸，DNA 就与聚合酶部分脱离（与小沟之间的氢键断裂但与拇指域之间的静电相互作用仍保持）。然后 DNA 用同样的序列非特异性机制，迅速地在移动一个碱基对的位置上与聚合酶重新结合。下面我们将要继续讨论，DNA 聚合酶与完全包围 DNA 的“滑动夹”蛋白之间的相互作用，可进一步提高延伸能力。

外切核酸酶校正新合成出的 DNA

仅依靠碱基配对的几何学以及碱基间互补性的系统，尚无法达到细胞内观测到的极高的 DNA 合成精确度（约每添加 10^{10} 个碱基对产生一个错误）。影响 DNA 聚合酶精确度的主要因素是碱基偶然地变换成“错误”的互变异构体（亚氨基或烯醇，图 6-5）。碱基的这种变构体使不正确的碱基对可以位于催化中正确的位置。如我们现在所述，校正过程可纠正这种错误。

DNA 合成的校正是通过核酸酶去除不正确碱基配对的核苷酸来实现。最初这种类型的酶被认作是 DNA 聚合酶的同一类多肽，现在则称为校正外切核酸酶（proofreading exonuclease）。这类酶可以从 DNA 的 3′端，即从新 DNA 链的生长端开始降解 DNA。（只能从 DNA 末端进行降解的酶叫做外切核酸酶，而能在 DNA 链中部切割的核酸酶叫做内切核酸酶。）

最初，3′外切核酸酶作为 DNA 聚合酶同一多肽的一个部分并没引起什么关注。为什么 DNA 聚合酶需要把它刚刚合成出的 DNA 降解掉呢？当了解到外切核酸酶有很强的降解含错误碱基对 DNA 的倾向时，它的功能就变得清晰了。这样，当发生不正确核苷酸被添加到引物链这种罕见的情况时，校正外切核酸酶可将此核苷酸从引物链的 3′端除去。这个新加长 DNA 的“校正”给了 DNA 聚合酶第二次添加正确核苷酸的机会。

DNA 聚合酶向不正确碱基配对的引物附近添加核苷酸的能力降低，从而加强了对错配核苷酸的去除。因为与手掌域相互作用差，所以错配 DNA 改变了 3′-OH 和引入核苷酸的几何构象。与添加不正确配对的 dNTP 使催化作用降低一样，改变的几何构象降低了核苷酸的添加速度。所以，当添加错配碱基时，它既降低了新核苷酸添加的速度，也增加了校正外切核酸酶的活性。

在 DNA 合成时，无需 DNA 脱离聚合酶即可进行校正（图 8-10）。当错配的碱基对存

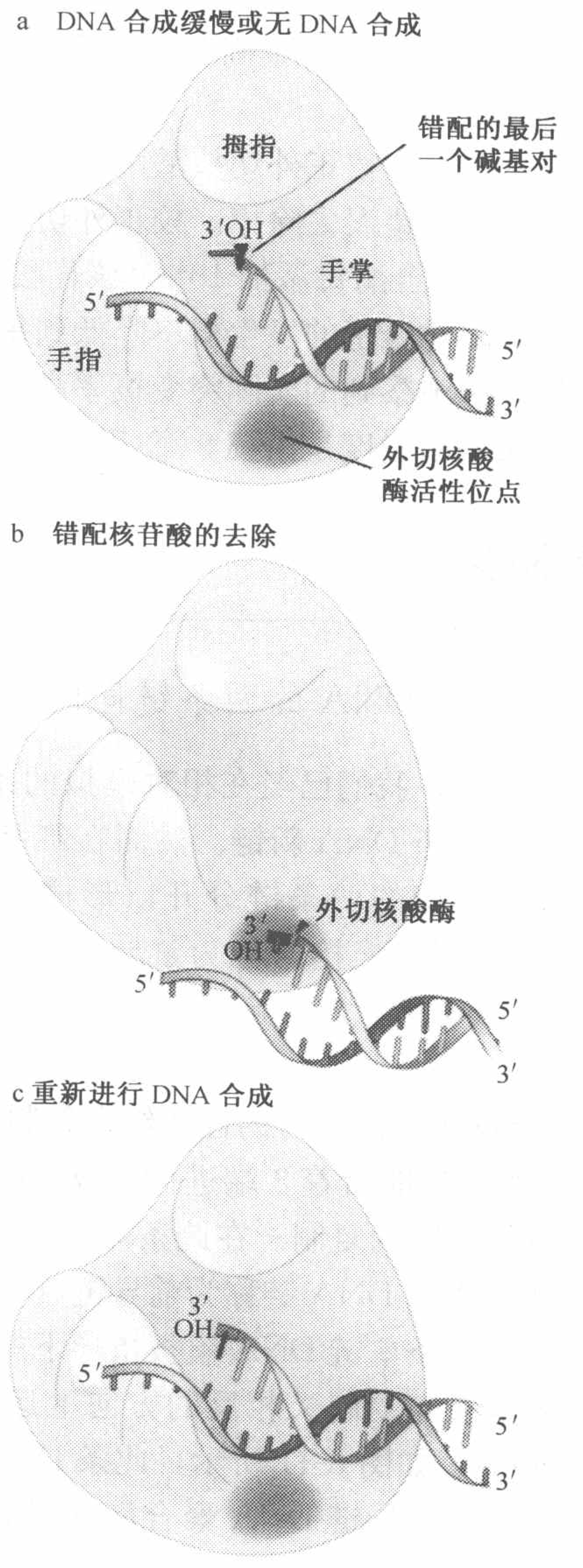


图 8-10 校正外切核酸酶将碱基从错配 DNA 的 3′端除去。 a. 当聚合酶把不正确核苷酸掺入 DNA 后，DNA 合成的速度降低且引物 3′端与 DNA 聚合酶活性位点的亲和力减弱。b. 当错配后，DNA 的 3′端与校正外切核酸酶活性位点的亲和力增加。与此活性位点结合后，错配核苷酸即被除去。c. 当错配核苷酸被除去后，正确配对的 DNA 与 DNA 聚合酶的亲和力恢复，DNA 合成继续进行。（来源：Adapted from Baker T. A. and Bell S. P. 1998. Polymerase and the replisome: Machines within machines. Cell 92:296, Fig. 1b. © Elsevier.）

在于 DNA 聚合酶的活性位点时，引物-模板接头会不稳固，产生若干不成对的碱基对 DNA，DNA 聚合酶对错配碱基的亲和力降低，但核酸外切酶在单链的 3′末端增强了 10 倍的活性因而新合成的未配对的 3′端由 DNA 聚合酶活性位点移到核酸外切酶活性位点，不正确的核苷酸被外切酶移除（其余的额外核苷酸也被移去）。当不正确碱基对被去除后，正确配对的引物-模板接头滑回 DNA 聚合酶的活性位点，DNA 合成继续进行。

本质上，校正外切核酸酶就像键盘上的“delete”（删除）键那样工作，它只能把最近发生的错误去除掉。校正外切核酸酶的加入极大地增加了 DNA 合成的精确度。平均 DNA 聚合酶每添加 10^5 个核苷酸就会插入 1 个不正确的核苷酸。校正外切核酸酶将不正确配对碱基的发生概率降低到每添加 10^7 个核苷酸出现 1 次。此错误概率仍比通常在细胞中观察到的实际突变概率高很多（每添加 10^{10} 个核苷酸出现 1 次错误）。这部分高出来的精确度是由第 9 章所述的复制后错配修复过程实现的。

复制叉

复制叉上 DNA 的两条链同时合成

至此，我们已经在相对模拟的条件下讨论了 DNA 的合成，即在引物-模板接头处仅产生一条 DNA 新链。然而在细胞中，DNA 双链的两条链是同时进行复制的。这要求将双螺旋的两条链分开以形成两条 DNA 模板。刚分开的模板链与未复制的双链 DNA 之间的连接区称为复制叉（replication fork）（图 8-11），复制叉向着未复制的 DNA 双链区域连续运动，其身后留下指导两个子代 DNA 双链形成的两条单链 DNA 模板。

DNA 反向平行的性质给在复制叉上两条暴露的模板同时进行复制制造了难题，因为 DNA 只能沿着 3′端进行合成，所以两条暴露的模板中仅有一条能够随着复制叉的运动进行连续的复制。在此条模板链上，DNA 聚合酶简单地尾随着复制叉。此模板指导下新合成的 DNA 链称为前导链（leading strand）。

另一条单链 DNA 模板指导下的新 DNA 链合成遇到的问题较多，此模板指导 DNA 聚合酶在与复制叉相反的方向上运动。此模板指导的新 DNA 链称为后随链（lagging strand）。如图 8-11 所示，此条 DNA 链必须以不连续方式进行合成。

虽然前导链 DNA 聚合酶在模板一解开后就能复制，但是后随链在其能够复制之前，必须等复制叉运动并暴露出足够长的模板后才能进行合成。当新的后随链模板暴露足够长度后，DNA 合成起始并持续至它抵达后随链 DNA 的上一个新合成片段的 5′端为止。

在后随链上形成的新链 DNA 短片段称为冈崎片段（Okazaki fragment），其在细菌中的长度变化为 1000～2000 个核苷酸，在真核生物中为 100～400 个核苷酸。冈崎片段被合成出来后，相互之间很快共价连接成一条连续的完整的新 DNA 链。因此，冈崎片段是 DNA 复制中短暂的中间产物。

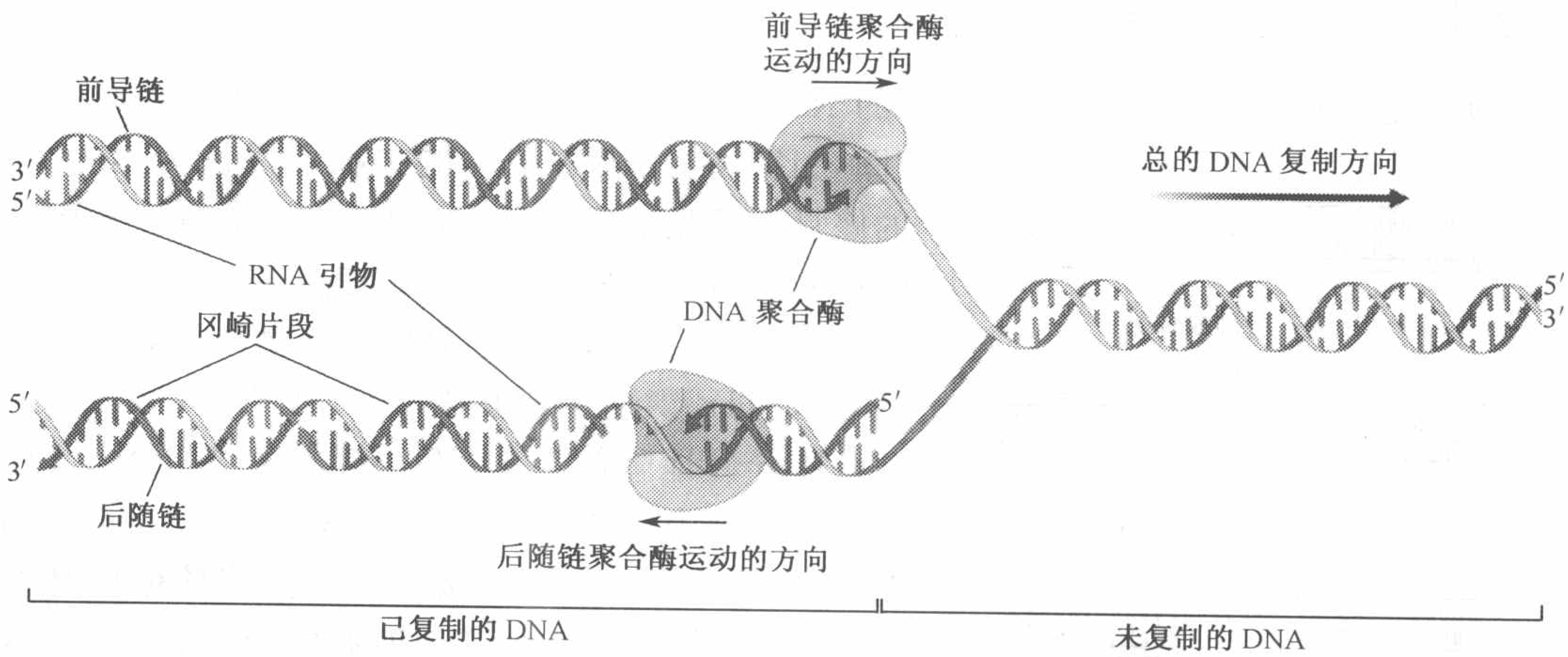


图 8-11 复制叉。新合成的 DNA 表示为红色，RNA 引物表示为绿色。为了方便说明，所示的冈崎片段人为地缩短了。在细胞中，冈崎片段在 100～1000bp 不等。

DNA 新链的起始需要一条 RNA 引物

如上所述，所有 DNA 聚合酶均需要带有游离 3′-OH 的引物，它们不能够从头启动新 DNA 链的合成。DNA 新链的合成是如何开始的呢？为了完成这一步，细胞利用了 RNA 聚合酶具有而 DNA 聚合酶没有的能力：从头开始合成新 RNA 链。引物酶（Primase）是一种能在单链 DNA 模板上制造短 RNA 引物（5～10 个核苷酸长度）的特殊的 RNA 聚合酶。这些引物随后由 DNA 聚合酶进行延伸。虽然 DNA 聚合酶只能将脱氧核糖核苷酸添加到 DNA 中，但它们用与 DNA 模板退火的 RNA 引物或 DNA 引物都能启动合成。

虽然前导链和后随链都需要引物酶来起始 DNA 的合成，但是引物酶在这两种链上工作的频率却相差极大（图 8-11）。每条前导链只需要一条 RNA 引物；相反，后随链的不连续合成意味着每个冈崎片段都需要新引物。因为一个复制叉可复制数百万个碱基对，所以后随链的合成需要数百至数千冈崎片段及其相关的 RNA 引物。

与 mRNA、rRNA 和 tRNA 合成中涉及的 RNA 聚合酶不同（第 12 章），引物酶不需要用特异 DNA 序列来起始新 RNA 引物的合成。

反而，引物酶偏好用含有一个特殊的三聚体的 ssRNA 模板来起始 RNA 的合成(在大肠杆菌引物酶中的为 GTA)。有意思的是在分析大肠杆菌基因组序列中发现大肠杆菌引物酶的 GTA 目标序列在基因组分中过量表达，将会是后滞链 DNA 合成的模板。

引物酶的活性将会明显增强当它和处于 DNA 复制叉的 DNA 解螺旋酶结合。此蛋白质在 DNA 复制叉处解开 DNA 链，产生可以被引物酶作用的 ssDNA 模板，DNA 解螺旋酶将在下面详细介绍。对 ssDNA 模板和另外一种复制叉蛋白的结合确保了引物酶只对复制叉处的解链 DNA 具有活性。

完成 DNA 复制必须除去 RNA 引物

要完成 DNA 的复制，用于起始的 RNA 引物就必须被除去并用 DNA 取而代之（图 8-12）。对 RNA 引物的去除可以认为是 DNA 的修复，且此步骤与 DNA 的切除修复有许多相同的性质，后者在第 9 章中有详细阐述。

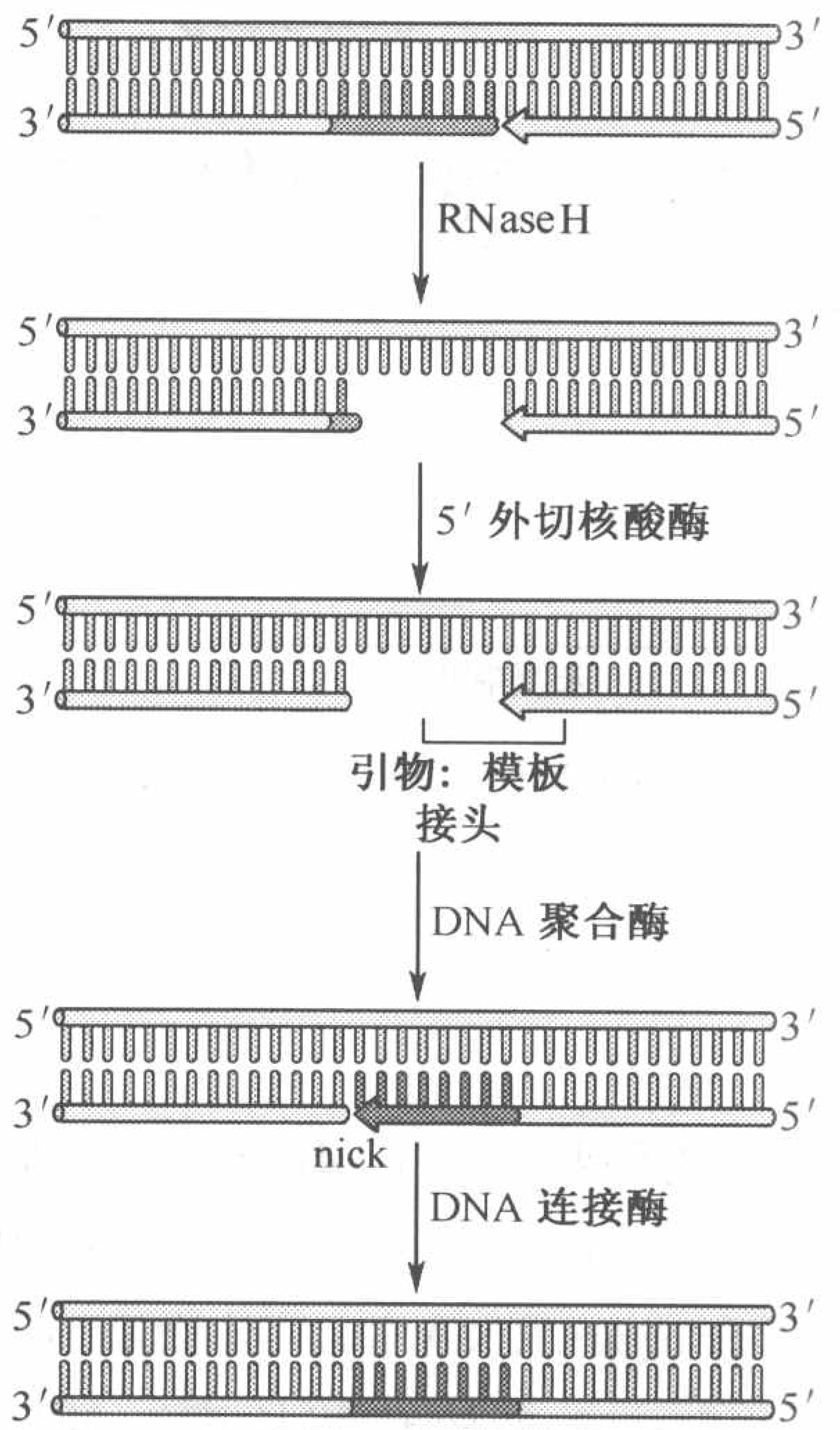


图 8-12 从新合成 DNA 上除去 RNA 引物。 图示为 RNA 引物去除过程中，RNase H、5′外切核酸酶、DNA 聚合酶以及 DNA 连接酶相继地作用。RNA 引物除去前的 DNA 显示为灰色，RNA 引物显示为绿色，新合成的替代 RNA 引物的 DNA 显示为红色。

为了用 DNA 取代 RNA 引物，RNA 酶 H（RNase H）识别并除去各条 RNA 引物的大部分。此酶特异性地降解与 DNA 碱基配对的 RNA（此处，其名称中的“H”意指 RNA-DNA 杂交链中的杂交）。除了与 DNA 末端直接连接的核糖核苷酸外，RNA 酶 H 可除去其他所有的 RNA 引物。这是因为 RNA 酶 H 只能断裂 2 个核糖核苷酸之间的键。最后一个核糖核苷酸是由从 5′端降解 RNA 或 DNA 的外切核酸酶除去的。

RNA 引物除去后在双链 DNA 中留下的缺口，形成引物-模板接头，这是 DNA 聚合酶理想的底物（图 8-12）。DNA 聚合酶填补此缺口直至每个核苷酸都碱基配对，留下一条完整的 DNA 分子（除了在骨架上被修复链的 3′-OH 和 5′磷酸基之间有一处断裂）。DNA 上的这个“切口”被称为 DNA 连接酶（DNA ligase）的酶修复。DNA 连接酶用一个高能协同因子（co-factor）（如 ATP），在毗邻的 5′磷酸基和 3′-OH 之间形成一个磷酸二酯键。只有当所有的 RNA 引物都被替代而且相关的切口均被填充后，DNA 的合成才结束。

DNA 解旋酶在复制叉之前解开双螺旋

一般而言，DNA 聚合酶解离双链 DNA 的两条碱基配对链的能力很差。因此，在复制叉上，由称为 DNA 解旋酶（DNA helicase）的另一组酶催化双链 DNA 两条链的分离。这些酶使用核苷三磷酸（一般为 ATP）水解的能量结合到单链 DNA 上，并沿着其定向移动，以去除任何退火到其所结合的单链 DNA 上的 DNA 链。在复制叉上作用的 DNA 解旋酶通常是环形的六聚体蛋白（图 8-13）。这些环形的蛋白质复合体环绕着单链-双链接头附近复制叉上的两条单链中的一条。

与 DNA 聚合酶一样，DNA 解旋酶也具有延伸性。在每次与底物结合后，就可解开 DNA 上的多个碱基对。在复制叉上发现的环形六聚体 DNA 解旋酶，因为环绕着 DNA 而显示有高度的延伸能力。因此，解旋酶从其 DNA 底物上释放下来需要打开六聚体蛋白环，这是非常罕见的。换句话说，解旋酶只有当它到达其环绕的 DNA 的末端

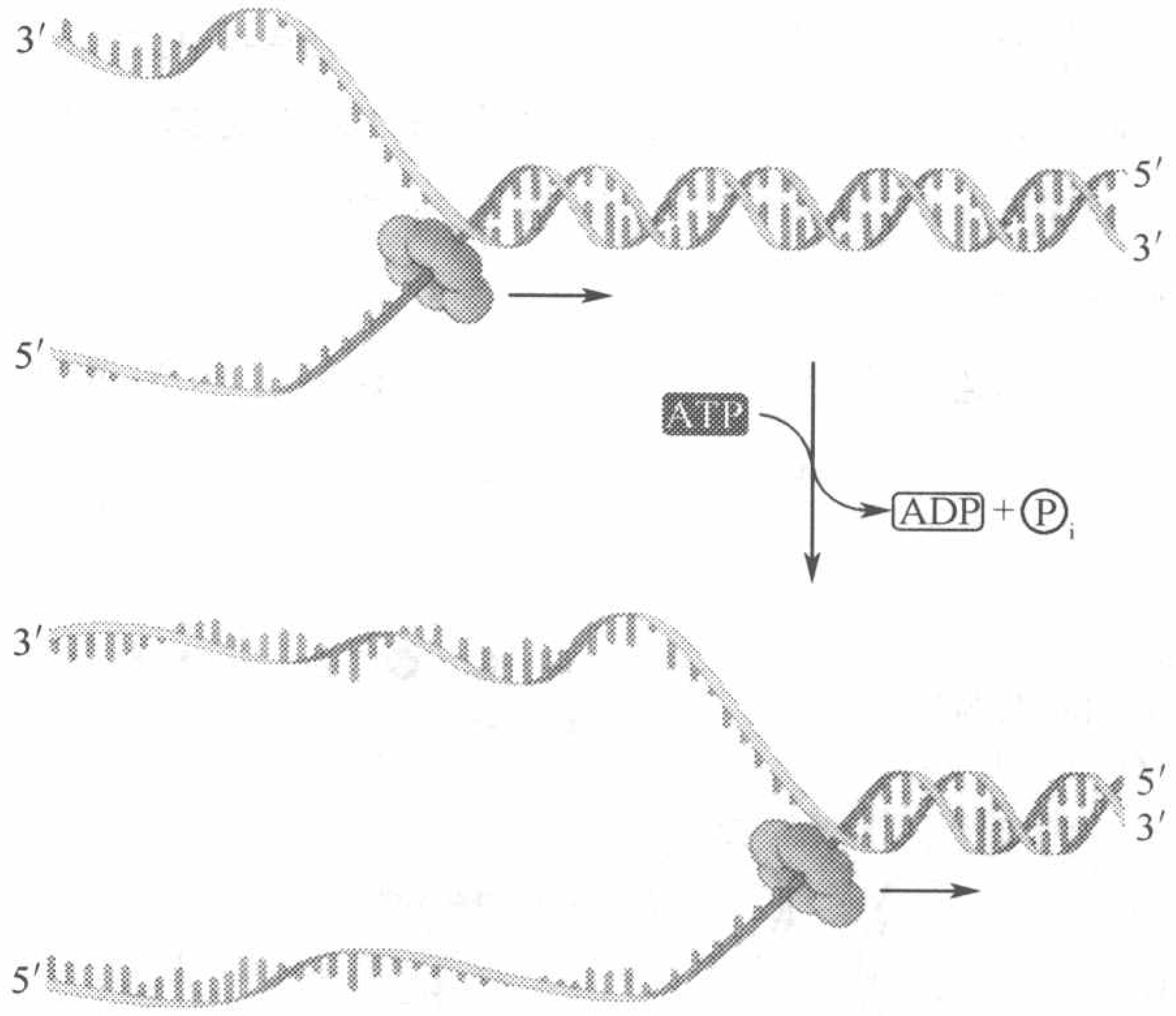


图 8-13 DNA 解旋酶分离双螺旋的两条链。当 ATP 加到与单链 DNA 结合的 DNA 解旋酶上时，解旋酶在单链 DNA 上以一定的极性运动。在图示例中，DNA 解旋酶具有5′→3′极性。这个极性说明 DNA 解旋酶与复制叉上的后随链模板结合。

时才解离下来。

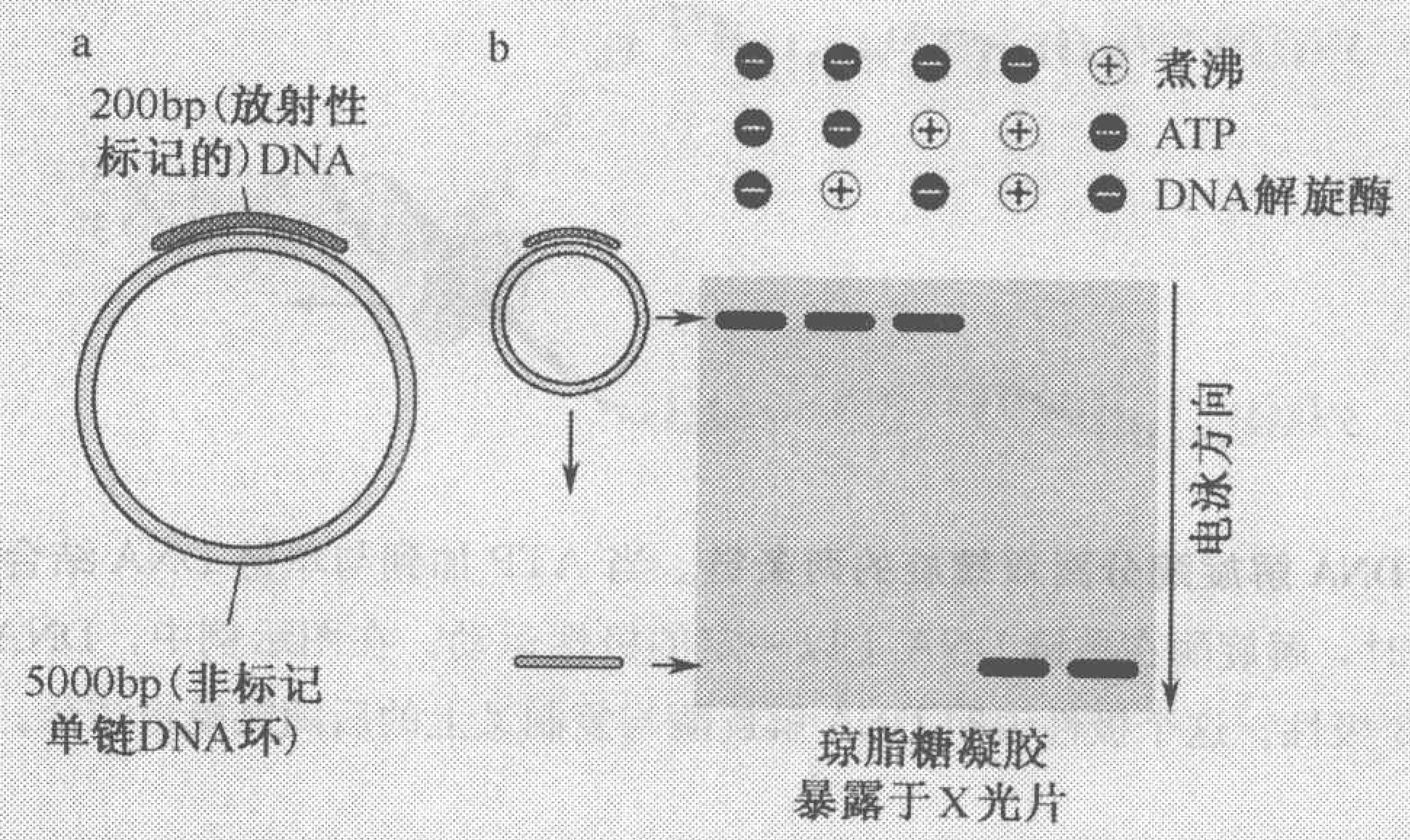
当然，酶和 DNA 的这种排列方式为最初 DNA 解旋酶与 DNA 底物的结合制造了麻烦。在环状染色体中没有 DNA 末端被 DNA 解旋酶穿过，这种问题也更为明显。因此，细胞中有特殊的机制来组装 DNA 解旋酶环绕 DNA（参见下面的“复制的起始”）。这种 DNA 复制中所涉及的蛋白质与其 DNA 底物之间的拓扑学连接是提高延伸能力的常见机制。

每个 DNA 解旋酶在单链 DNA 上都沿着一定的方向运动。这是所有 DNA 解旋酶都具有的特性，称作极性（polarity）（框 8-3）。DNA 解旋酶可以有5′→3′或3′→5′的极性，此方向始终是根据结合的（或被环形解旋酶环绕的）DNA 链而不是去除的那一条链决定的。以在复制叉的后随链模板上作用的 DNA 解旋酶为例，其5′→3′的极性使 DNA 解旋酶朝复制叉双链区域行进（图 8-13）。与所有以定向方式沿 DNA 运动的酶一样，解旋酶沿单链 DNA 的运动也需要化学能量的提供。解旋酶的能量由 ATP 水解提供。

框 8-3 DNA 解旋酶极性的确定

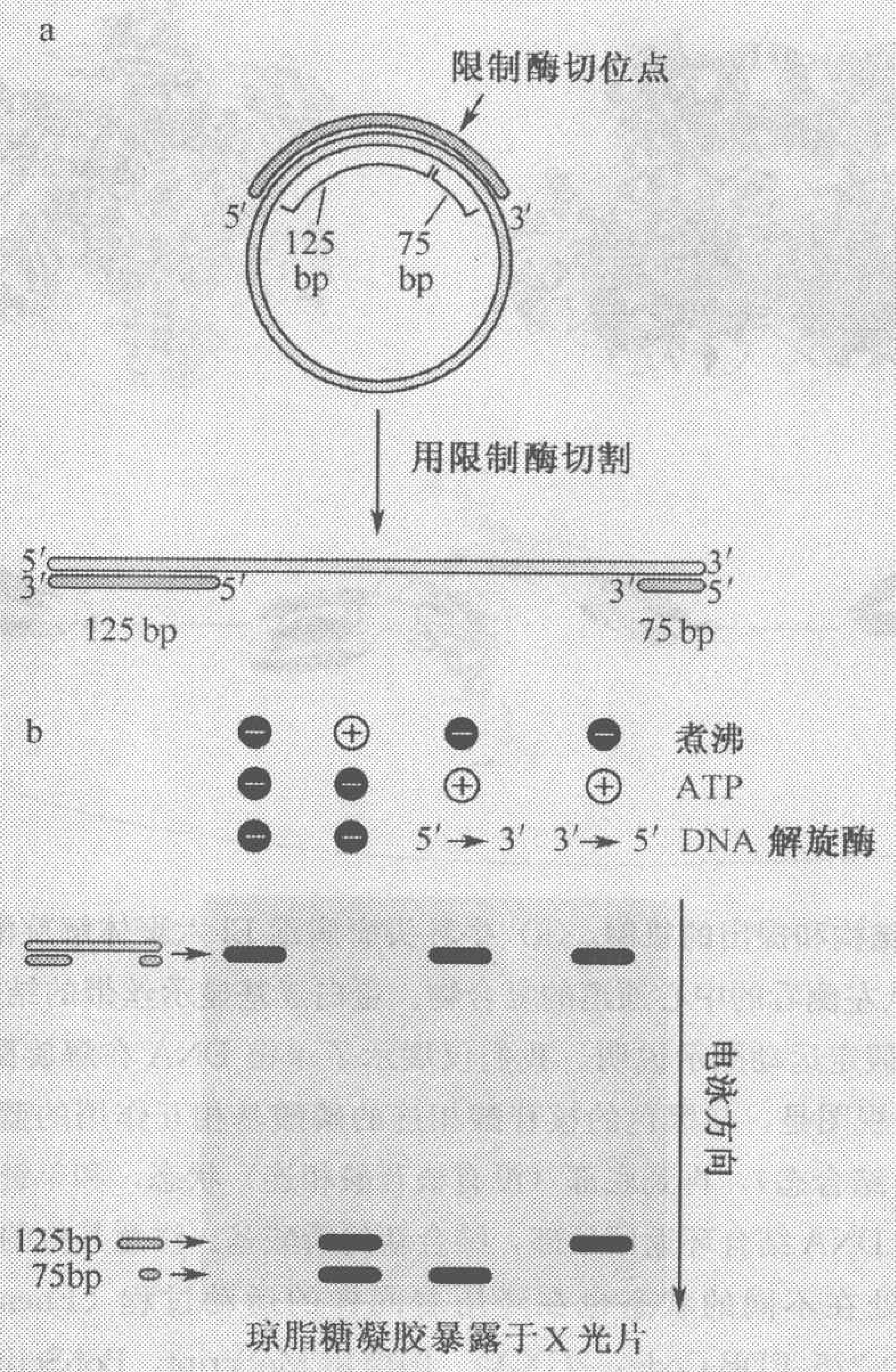
DNA 解旋酶的活性可通过其将 DNA 双链的一条链与另一条链分离的能力来测定。在典型的 DNA 解旋酶实验中，底物由一条短的、已标记的单链 DNA 构成，它与一条长的、未标记的单链 DNA 退火（通常，标记是将放射性^{32}P 掺入到短的单链 DNA 中）。例如，大的环形单链 DNA（如 5000 个碱基）与短的（如 200 个碱基）标

记线性单链DNA分子杂交（框8-3图1）。DNA解旋酶将把短的线性单链DNA从大的单链DNA环上解下。链的分离可通过在非变性琼脂糖凝胶上短的标记单链DNA电泳迁移率的变化来测定（第20章）。凝胶暴露于只能检测到放射性标记DNA的X光片上以后，可检测到短的DNA在凝胶上的位置。当它与单链DNA环杂交的时候，将与大的单链DNA环一起迁移。相反，当短的单链DNA被DNA解旋酶从单链DNA环上解下后，它将按照其200bp的实际大小来迁移。



框8-3图1 DNA解旋酶活性的生物化学实验。 a. 检测解旋酶活性的DNA底物。一条5000 bp非标记单链环形DNA与一条200bp的放射性标记DNA退火。为了方便起见，两个分子没有按比例绘制。b. 为了检测DNA解旋酶的活性，DNA底物与DNA解旋酶接触（在有ATP和没有ATP的情况下）。反应后，产生的DNA分子用琼脂糖凝胶电泳（非变性）进行分离。当短的放射性标记的DNA与大的单链DNA环碱基配对的时候，两个分子作为一个大分子共同迁移；相反，DNA解旋酶作用后，短的放射性标记单链DNA迁移的位置与其长度一致。琼脂糖凝胶暴露于X光片之后，只能看到放射标记的DNA位置。作为对照，两个DNA分子可通过煮沸分离。同样也能使碱基配对区域变性。

将这个简单的实验修饰后可用于确定DNA解旋酶的极性。假设有一个限制酶的切割位点位于碱基配对区域内的不对称位置上（框8-1图2）。当这个位点被切割后，将产生一个大的单链线性DNA，其两端有两个长度不等的双链DNA区域。注意DNA解旋酶与单链DNA而不是与双链DNA结合。所以，DNA解旋酶唯一能与此新线性底物结合的部位位于两个双链DNA区域之间。因为DNA解旋酶的极性，任何特定的DNA解旋酶都只能把两个短的单链DNA中的一个解下。因为两个短的单链DNA的长度不等，所以解离片段的大小将显示DNA解旋酶按哪个方向沿着线性底物的单链DNA区域运动。



框 8-3 图 2　DNA 解旋酶极性的生物化学实验。 a. DNA 底物。与图 1 所示的相同 DNA 底物用产生平末端的限制酶切割。选择不对称切割的限制酶，产生与 5000 bp 未标记单链 DNA 端退火的 125bp 和 75bp 的放射性标记单链 DNA 片段。产生的 DNA 分子 5′和 3′端都被标出。b. 图示为 X 光片暴露用来分离 DNA 解旋酶处理后 DNA 产物的琼脂糖凝胶。在 a 部分产生的底物可与 DNA 解旋酶孵育，以此确定解旋酶的极性。图示为5′→3′和3′→5′ DNA 解旋酶的结果。底物的煮沸显示了所有碱基配对都彻底变性的结果。

DNA 解螺旋酶牵拉单链 DNA 通过蛋白质中心的孔

一个六聚体 DNA 解螺旋酶如何利用 ATP 水解酶释放的能量移动 DNA？一种病毒的六聚体解螺旋酶结合到单链 DNA 的原子结构提供了这个问题的突破点。在这种结构中，单链 DNA 被解旋酶的六个亚基所环绕（图 8-14）。每个亚基具有一个“发夹”蛋白环结合到 DNA 主干的磷酸基和它两个邻近的核糖组分。有意思的是，这些 DNA 结合的环发现是右手螺旋，沿着单链 DNA 每一个环结合到下一个磷酸基团。如图 8-14 所示，最顶端的“阶梯”结合到 5′端，最低层结合到单链 DNA 的 3′端。

这张原子结构图只是一个快照；但是，六个不同的亚基在 DNA 异位过程中处于不同的阶段。结合六个不同亚基同 DNA、ATP/ADP 的相互作用表明了蛋白质的发夹结构如何协同运动来牵拉单链 DNA 通过解旋酶的中心孔。一个亚基首先结合到单链 DNA 结构的头部分（图 8-14b），DNA 结合环通过连续的构象移向底部，拉着底部的

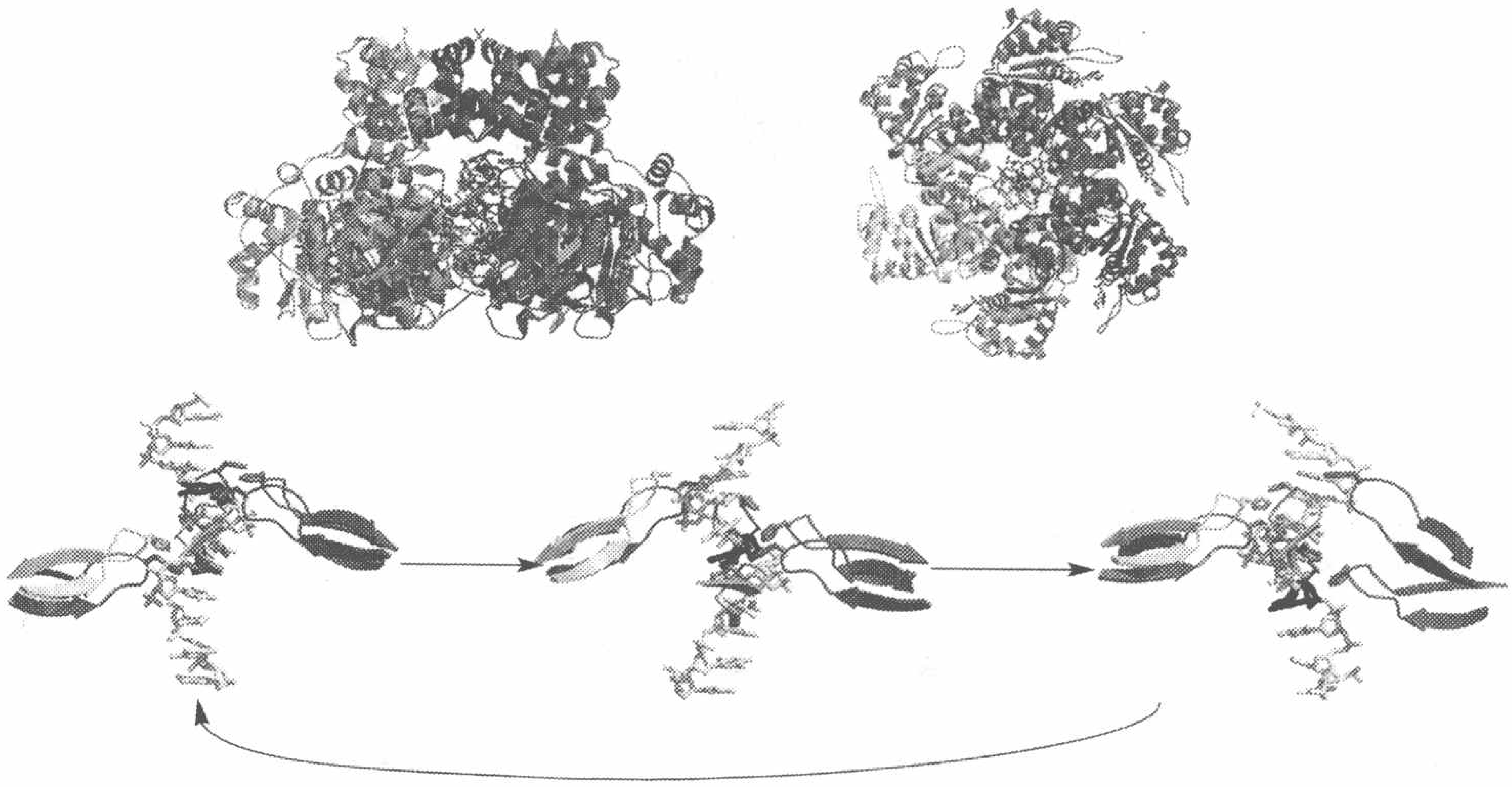


图 8-14 DNA 解旋酶的结构和提出的机制。(a) 牛乳头瘤病毒 El 六聚体解旋酶的整体结构。每个亚基显示出了不同的颜色，从左侧看的中心通道的复合物。蛋白亚基显示丝带的结构，单链 DNA 用棒图显示。(b) DNA 结合发夹假定运动图示说明。我们仅展示了单链 DNA 在解旋酶的中心通道，关键的发夹结合到 DNA 上。三个视图是，与黑色的核苷酸相连的磷酸基相互作用的蓝色发夹从顶端（ATP 结合）状态到中间（ADP 结合态），再到底部（没有核苷酸相连）状态，和单链 DNA 随着发夹移动。重新结合 ATP 认为是驱动 DNA 结合环返回顶部，结合新的磷酸基。注意其他的五个 DNA 结合环显示了不同的颜色，说明它们处在不同的状态也在经历着同样的运动过程（Enemark E. J. and Joshua-Tor L. 2006. Nature 442：270-275. PDB Code：2GXA）。图像用 MoScript、BobScript 和 Raster 3D 处理。

DNA。重要的是，不同的构象结合着不同的核苷酸状态；顶部是 ATP 结合状态，中间是 ADP 结合状态，底部缺少结合的核苷酸。因此，每一个单一的亚基结合点、水解酶、游离的 ATP，都将经历从顶部、中间、底部的构象。总之可以把解旋酶想象成为有六只手，六只手交互地拉一根绳子的样子。

除了单链的转位，解旋酶必须置换补偿链来使 DNA 解旋。六聚解旋酶的中心通道结构表明，链的置换是为了使 DNA 通过中心通道. 中心通道最窄的部位直径是 13Å，对单链 DNA 来说足够大，但是对直径为 20Å 的双链 DNA 而言太小了。

复制前单链结合蛋白使单链 DNA 稳定

DNA 解旋酶经过之后，新产生的单链 DNA 必须保持碱基未配对的状态，直至可被用作 DNA 合成的模板为止。为了使分开的链稳定，单链 DNA 结合蛋白（简称为 SSB）迅速地结合其上。一个 SSB 的结合会促进另一个 SSB 与其紧邻的单链 DNA 结合（图 8-15）。这称为协同结合（cooperative binding)，其发生是因为与紧邻的单链 DNA 区域结合的 SSB 分子相互之间也能够结合。这样大大稳定了 SSB 与单链 DNA 之间的相互作用，使得已有一个或多个 SSB 结合的位点比其他位点更易结合新的 SSB。

协同结合使单链 DNA 随着 DNA 解旋酶的释放而很快被 SSB 所覆盖（协同结合是许多 DNA 结合蛋白的一种特性，见框 16-4)。一旦被 SSB 覆盖，单链 DNA 即处于伸

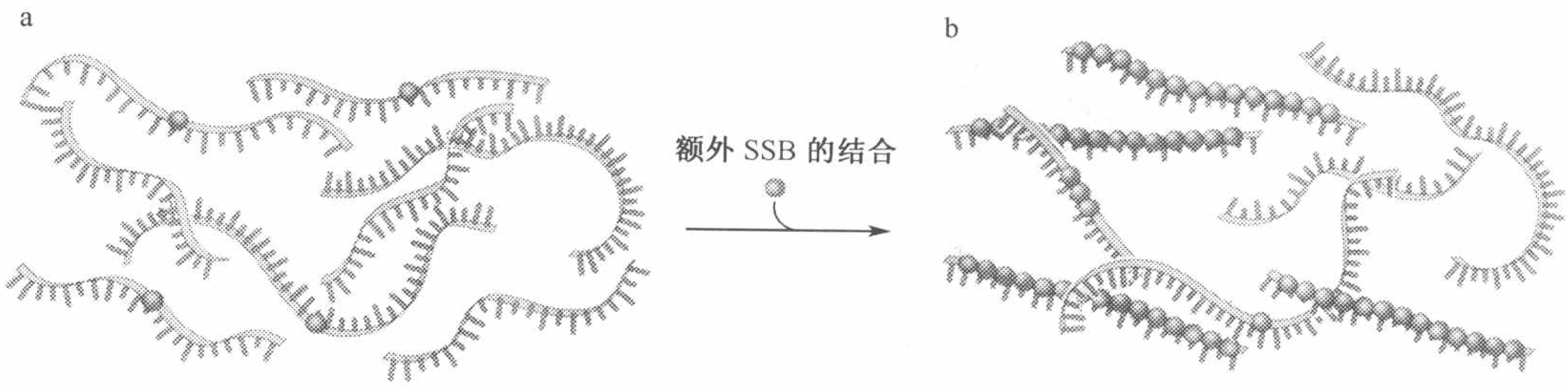


图 8-15 单链 DNA 结合蛋白（SSB）与 DNA 的结合。 a. 图中显示有限量的 SSB 与 9 条单链 DNA 分子中的 4 条结合。b. 当更多的 SSB 与 DNA 结合后，它优先结合到先前已有 SSB 分子结合的 DNA 附近。只有最初结合的单链 DNA 被 SSB 完全覆盖后，其他 DNA 分子才发生结合。注意当单链 DNA 被 SSB 覆盖后，变成更为伸直的构象，以抑制分子间的碱基配对。

直状态，有利于其作为模板进行 DNA 合成或 RNA 引物的合成。

SSB 以序列非特异性方式与单链 DNA 相互作用。SSB 主要通过与磷酸骨架的静电相互作用以及与 DNA 碱基的相互堆积作用与单链 DNA 接触。与序列特异性的 DNA 结合蛋白不同，SSB 与单链 DNA 碱基之间几乎没有氢键的作用。

拓扑异构酶除去 DNA 解螺旋时在复制叉上产生的超螺旋

随着复制叉上 DNA 链的分离，叉前的双链 DNA 变得更加正超螺旋化（图 8-16)。这种超螺旋的积累是 DNA 解旋酶消除了两链间碱基部分的结果。如果 DNA 链保持不断裂，则不会有调节 DNA 双链解旋的连环数（两条 DNA 互绕的次数）的减少（第 6 章)。因此，随着 DNA 解旋酶的行进，DNA 必须在越来越少的碱基对中维持同样数目的连环数。事实上，当超螺旋数保持不变时，每 10 个 DNA 碱基解旋时就有约 1 个 DNA 连环被除去。如果没有机制来消除这些超螺旋累积的话，复制机器将在不断上升的压力面前陷于停顿。

对于细菌的环形染色体此问题非常明显，但真核生物染色体也有同样的问题。因为真核生物染色体不是闭合的环，原则上它们能通过沿着其长度方向的旋转来消除引入的超螺旋。然而事实并非如此，因为不可能每次当螺旋解开一圈时，就旋转长达几百万碱基对长度的 DNA。

由 DNA 解旋酶的作用所引入的超螺旋，可通过拓扑异构酶对在复制叉前未被复制的双链 DNA 的作用来消除（图 8-16)。这些酶是通过断裂一条或两条 DNA 链但不让其脱落，并且让同等数目的 DNA 链通过断口来完成的（如我们在第 6 章所述)。此作用消除了超螺旋的累积。通过这种方式，拓扑异构酶作为一个“转环”，快速地消除了 DNA 解旋所产生的超螺旋累积。

复制叉酶扩大了 DNA 聚合酶底物的范围

DNA 聚合酶本身只能使与单链 DNA 模板退火的 3′-OH 引物有效地延长，引物酶、DNA 解旋酶以及拓扑异构酶的加入极大地扩大了 DNA 聚合酶可能的底物范围。引物

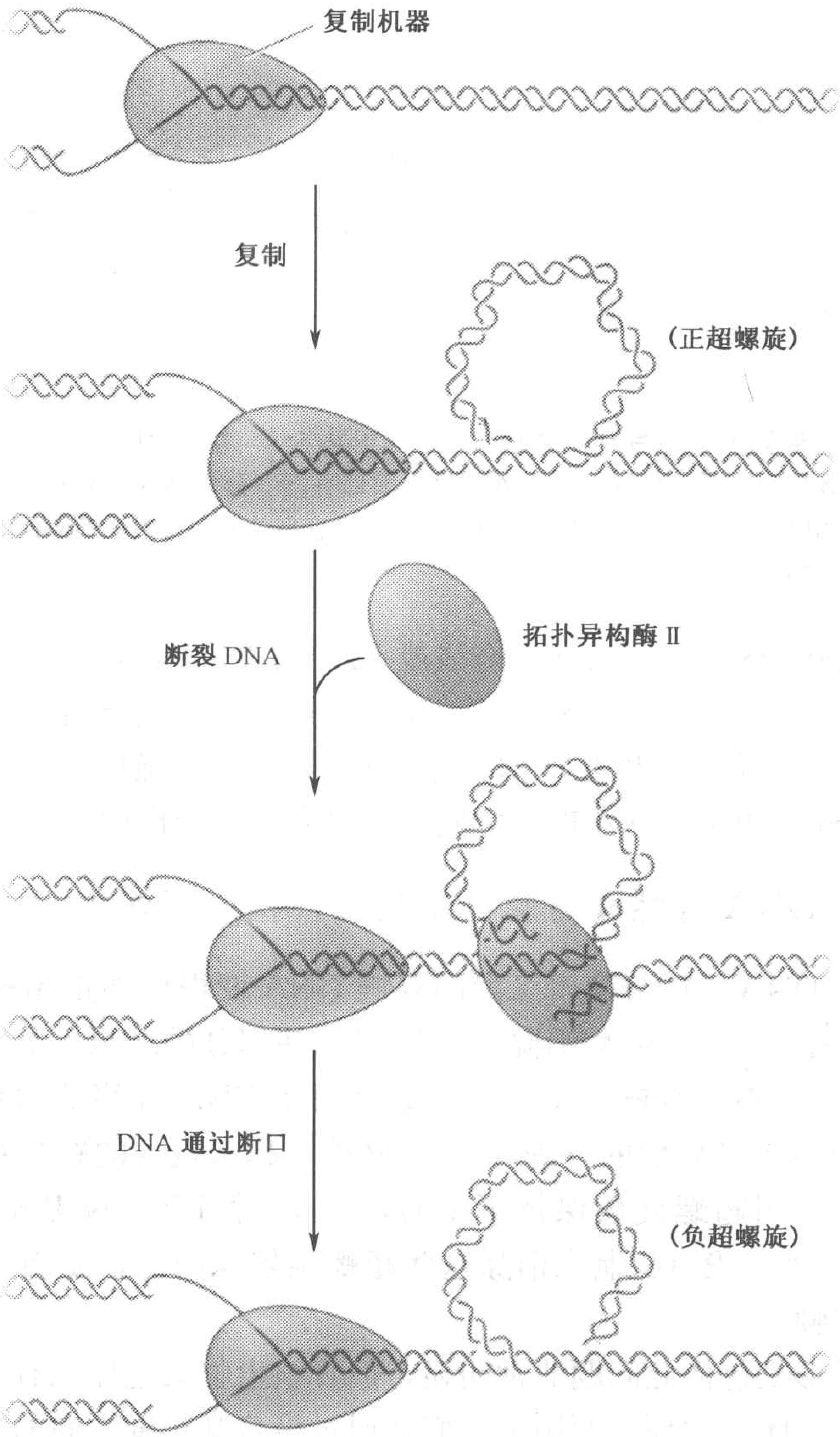


图 8-16 拓扑异构酶在复制叉上的作用。当正超螺旋在复制叉前累积时，拓扑异构酶会将其迅速地除去。本图显示拓扑异构酶Ⅱ除去由复制叉所产生的正超螺旋，将未复制的双链 DNA 中的一部分通过位于未复制区域附近的双链断口可以消除正超螺旋。值得注意的是，这种变化将使连环数减少 2 个，因此每复制 20bp 会发生一次。虽然此处所示的是Ⅱ型拓扑异构酶的作用，但是Ⅰ型拓扑异构酶也能消除复制叉产生的正超螺旋。

酶提供了在任何单链 DNA 上都能起始新的 DNA 链的能力。当然，引物酶的使用也提出了要除去 RNA 引物来完成复制的要求。类似地，DNA 解旋酶对链的解离以及拓扑异构酶对正超螺旋的消除使 DNA 聚合酶能够复制双链 DNA。虽然不同生物中这些蛋白质的名称不同（表 8-1），但是，细菌、酵母和人等生物均使用这一套酶的活性来完成染色体 DNA 的复制。

表 8-1　复制叉上工作的酶

	大肠杆菌（*E. coli*）	酿酒酵母（*S. cerevisiae*）	人类
引物酶	DnaG	引物酶（PRI1/PRI2）	引物酶
DNA 解旋酶	DnaB	Mcm 复合体	Mcm 复合体
SSB	SSB	RPA	RPA
拓扑异构酶	旋转酶、拓扑异构酶Ⅰ	拓扑异构酶Ⅰ、Ⅱ	拓扑异构酶Ⅰ、Ⅱ

值得注意的是，DNA 解旋酶和拓扑异构酶行使其功能时，均没有永久性地改变 DNA 的化学结构或合成任何新的 DNA 分子。DNA 解旋酶仅仅把将 DNA 双链维持在一起的氢键打开而没有断裂任何共价键。虽然拓扑异构酶断裂掉一个或多个 DNA 的共价键，但在 DNA 脱离之前每个断裂的键都被精确地修复了（图 6-25）。这些酶的作用是使 DNA 分子的构象发生改变，而不是改变 DNA 的化学结构。重要的是，这些构象的变化对于作为细菌和真核生物染色体基础的大的双链 DNA 分子的复制是至关重要的。

在复制叉上工作的蛋白质与 DNA 密切作用，但采用的是序列非特异性的方式。这些相互作用利用的是 DNA 与特定碱基对无关的特性：负电荷以及磷酸骨架结构（如 DNA 聚合酶的拇指域）、小沟内残基的氢键（如 DNA 聚合酶的手掌域）、碱基之间的疏水堆积作用（如 SSB）。另外，很多蛋白质具有环绕（如 DNA 解旋酶）或包裹（如 DNA 聚合酶）DNA 的结构，以保持与 DNA 的结合。

DNA 聚合酶的特化

细胞中 DNA 聚合酶为行使不同的功能而被特化

DNA 聚合酶在高效而精确地复制基因组过程中所处的核心地位要求细胞有多种被特化的 DNA 聚合酶。例如，*E. coli* 中至少有 5 种 DNA 聚合酶，这些酶依据其酶学特性、亚基组成和丰度而划分（表 8-2）。DNA 聚合酶Ⅲ（DNA PolⅢ）是染色体复制中所涉及的最主要的酶。因为全长为 4.6 Mb 的 *E. coli* 由 2 个复制叉进行复制，DNA 聚合酶Ⅲ必须有很高的延伸能力。与这些要求相一致的是，人们发现 DNA 聚合酶Ⅲ常常作为更大的具有高度延伸能力的复合体——DNA 聚合酶Ⅲ全酶（DNA PolⅢ holoenzyme）的一部分。

表 8-2　DNA 聚合酶的活性和功能

	亚基数目	功　能
原核生物（*E. coli*）		
PolⅠ	1	RNA 引物的去除、DNA 修复
PolⅡ（Din A）	1	DNA 修复
PolⅢ核心酶	3	染色体复制

续表

	亚基数目	功　能
Pol Ⅲ全酶	9	染色体复制
Pol Ⅳ（Din B）	1	DNA 修复、移损合成（trans lesion synthesis）（TLS）
Pol V（UmuC、$UmuD'_2C$）	3	移损合成
真核生物		
Polα	4	DNA 复制过程中合成引物
Polβ	1	碱基切除修复
Polγ	3	线粒体 DNA 复制及修复
Polδ	2～3	DNA 复制、核苷酸和碱基切除修复
Polε	4	DNA 复制、核苷酸和碱基切除修复
Polθ	1	DNA 交联修复
Polξ	1	移损合成
Polλ	1	减数分裂相关的 DNA 修复
Polμ	1	体细胞超变
Polκ	1	移损合成
Polη	1	相对精确的移损合成 past 顺-反环丁烷二联体
Polι	1	移损合成、体细胞超变
Rev1	1	移损合成

数据源自 Sutton 和 Walker，2001 及其参考文献。

相反，DNA 聚合酶Ⅰ（DNA Pol Ⅰ）专门用于除去起始 DNA 合成所需的 RNA 引物。为此，DNA 聚合酶Ⅰ具有 5′外切核酸酶的活性，能将 DNA 合成位点最上游的 RNA 或 DNA 迅速除去。与 DNA 聚合酶Ⅲ不同，DNA 聚合酶Ⅰ的延伸能力不强，每次结合仅能添加 20～100 个核苷酸。这些性质使其适用于 RNA 引物的去除和在所产生的单链 DNA 缺口中进行 DNA 的合成。DNA 聚合酶Ⅰ的 5′外切核酸酶可以除去 RNA 酶 H 所不能除去的 RNA-DNA 连接（图 8-12）。DNA 聚合酶Ⅰ合成较短片段的能力使其适用于合成先前被 RNA 引物占据的短区域（<10 个核苷酸）。

因为 DNA 聚合酶Ⅰ和 DNA 聚合酶Ⅲ都与 DNA 复制有关，所以这两种酶都必须有高度的精确性。因此，这两种蛋白质都携带有相关的校正外切核酸酶。*E. coli* 中的另外 3 种 DNA 聚合酶特别用于 DNA 的修复，它们无校正活性，这些酶将在第 9 章中予以讨论。

真核细胞中也有多种 DNA 聚合酶，通常细胞中有 15 种以上。其中，对于基因组的复制至关重要的有 3 种：DNA 聚合酶 δ、DNA 聚合酶 ε 以及 DNA 聚合酶 α/引物酶。这些真核生物的 DNA 聚合酶由多个亚基构成（表 8-2）。DNA Pol α/引物酶特别参与新 DNA 链的起始，这种四亚基蛋白复合体由二亚基的 DNA 聚合酶 α 和二亚基的引物酶组成。引物酶合成 RNA 引物后，所产生的 RNA 引物-模板接头立即与 DNA 聚合酶 α 结合以启动 DNA 的合成。

因为 DNA Pol α/引物酶的延伸能力相对较低，所以很快就被高延伸性的 DNA 聚

合酶 δ 或聚合酶 ε 取代。聚合酶 δ 或聚合酶 ε 取代 DNA Pol α/引物酶的过程称作聚合酶的切换（polymerase switching）（图 8-17），这导致在真核细胞复制叉上有 3 种不同的 DNA 聚合酶在工作。如同在细菌细胞中那样，真核细胞中其他的 DNA 聚合酶大多参与 DNA 的修复。

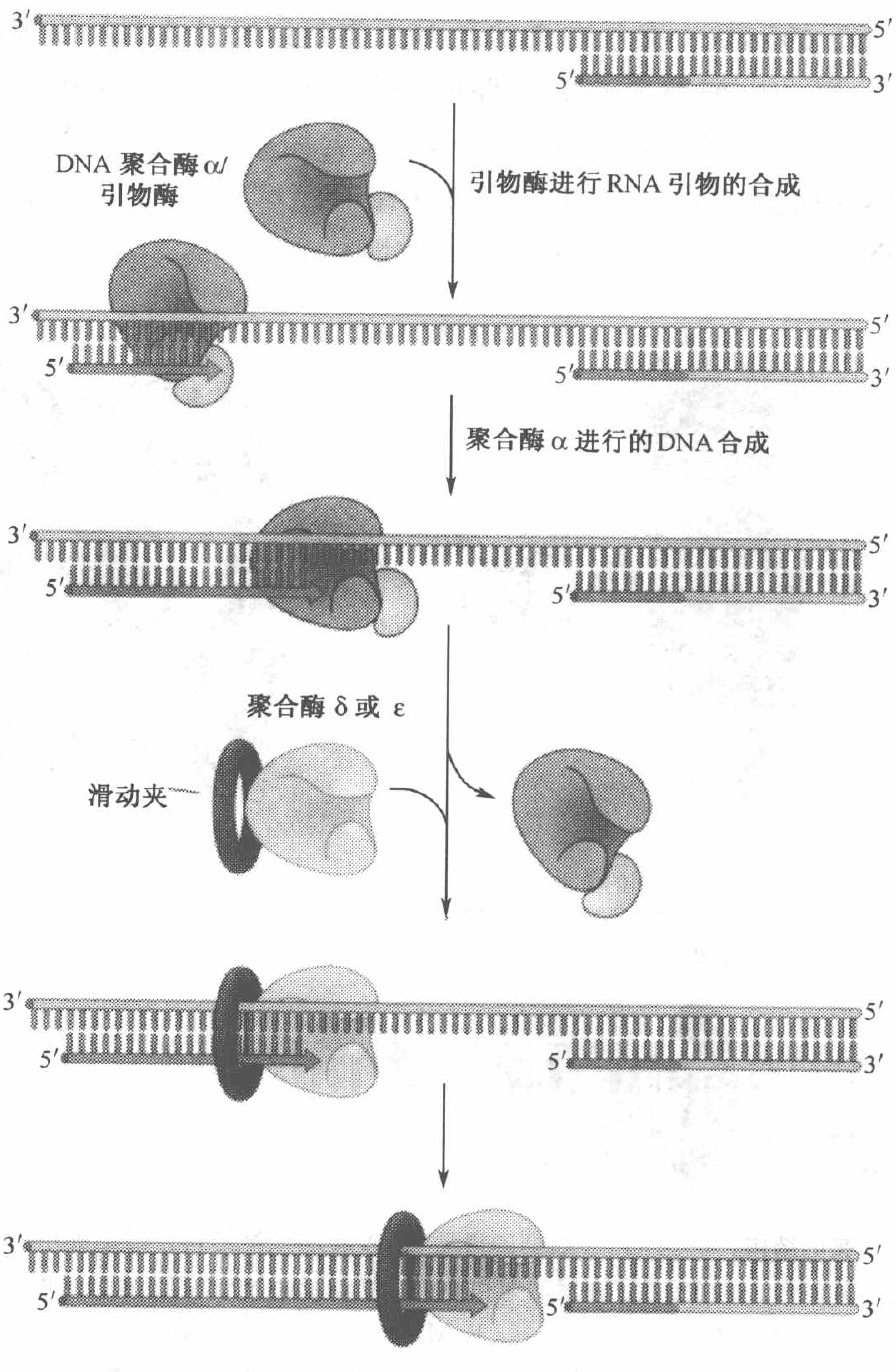


图 8-17　真核细胞 DNA 复制过程中 DNA 聚合酶的切换。为了图示方便，DNA 合成的长度比实际的短。通常 DNA 聚合酶 α/引物酶的产物为 50～100bp，由聚合酶 δ 或聚合酶 ε 进一步延伸 100～10 000 个核苷酸。虽然 DNA 聚合酶 δ 和 DNA 聚合酶 ε 都能替代 DNA 聚合酶 α/引物酶，但它们似乎是在特定 DNA 链（前导链和后随链）的复制中起作用。目前仍在对哪个聚合酶在哪条链上作用进行研究。

滑动夹极大地增加了 DNA 聚合酶的延伸能力

复制叉上高度的延伸能力确保了染色体的快速复制。如我们所讨论的，DNA 聚合酶

在复制叉上可以在不脱离模板的情况下合成成千上万的碱基对。尽管如此，在没有其他蛋白质存在的情况下，DNA 聚合酶在复制叉上仅能合成 20～100 个碱基对后即从模板上脱离下来。那么，是什么原因使这些酶在复制叉上的延伸能力如此显著的提高了呢？

DNA 聚合酶在复制叉上具有高延伸能力的关键原因在于它们与被称为滑动 DNA 夹（sliding DNA clamp）的蛋白质间的结合。这些蛋白质由多个相同的亚基构成，并组装成“油炸圈饼”的形状。夹子中央的孔洞大得足以环绕 DNA 双螺旋并在 DNA 和蛋白质之间留下一层一个或两个水分子厚的空间（图 8-18a）。这些夹子沿着 DNA 滑动，并不与 DNA 分离，在复制叉处还与 DNA 聚合酶紧密结合。这样，夹子环绕着新合成出的双链 DNA 而聚合酶与引物-模板接头结合（图 8-18b）。聚合酶与滑动夹所形成的复合体在 DNA 合成时沿着 DNA 模板高效地移动。

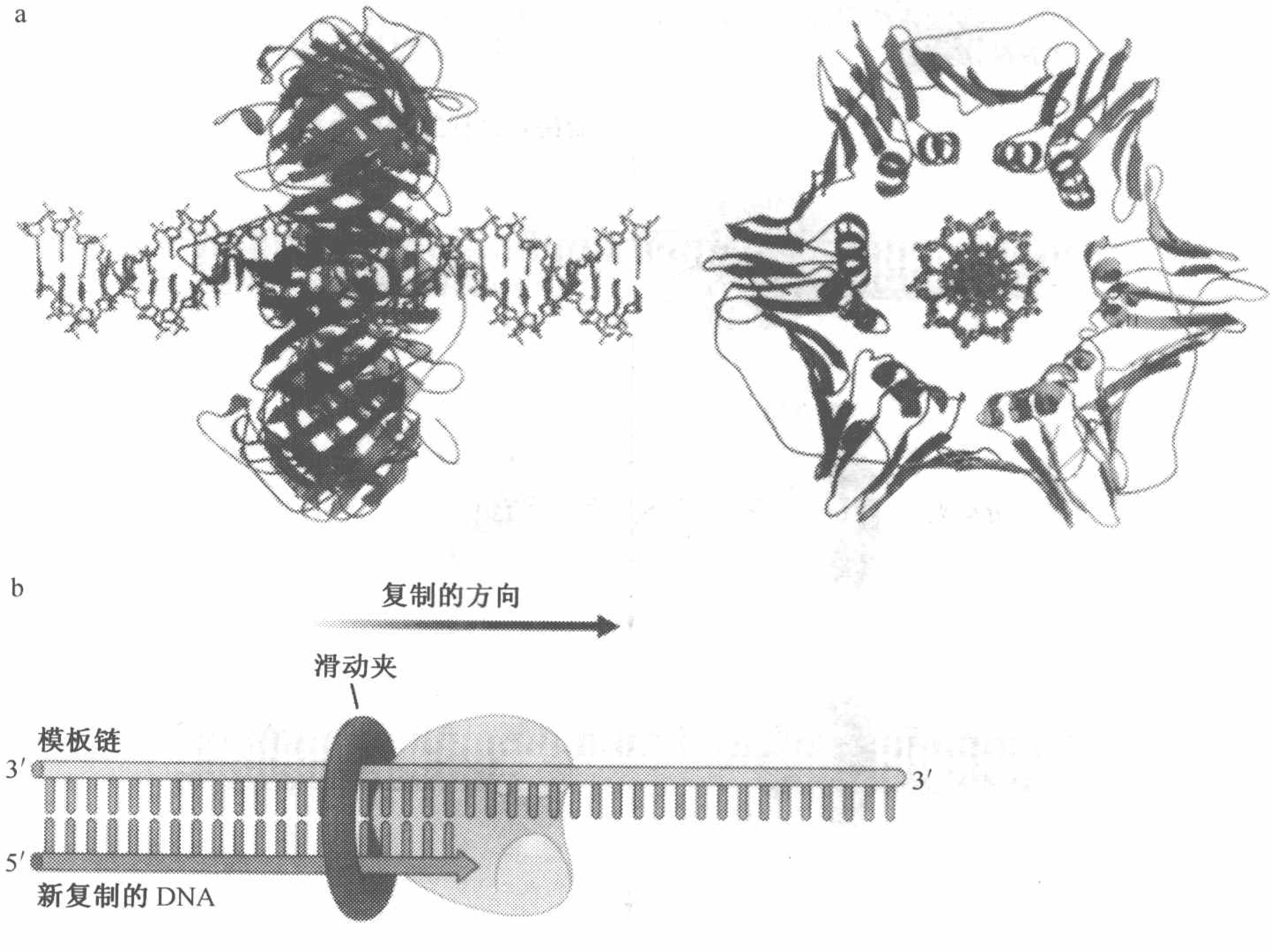


图 8-18 滑动 DNA 夹的结构。 a. 与 DNA 结合的滑动 DNA 夹的三维结构。滑动夹中央的开孔约为 35Å，DNA 螺旋的宽度约为 20Å，这样提供了在滑动夹和 DNA 之间容纳一个或两个水分子厚度薄层的足够空间。(Krishna T. S.，Kong X. P.，Gary S.，Burgers，P. M.，and Kuriyan J. 1994. Cell 79：1233.）图像用 Bob-Script、MolScript 和 Raster 3D 绘制。b. 滑动夹环绕着由与其结合的 DNA 聚合酶所产生的新复制的 DNA。当新合成的 DNA 从 DNA 聚合酶上出现时，滑动夹与最靠近新合成 DNA 的 DNA 聚合酶部分相互作用。

与滑动夹的结合是如何改变 DNA 聚合酶的延伸能力的呢？在无滑动夹的情况下，DNA 聚合酶平均每合成 20～100 个碱基对就从 DNA 模板上解离并散落开来。在有滑动夹的情况下，DNA 聚合酶的活性部位仍经常从 DNA 的 3′-OH 末端脱落，但是其与滑动夹的结合阻止了聚合酶从 DNA 上散开（图 8-19）。通过保持 DNA 聚合酶与 DNA 的紧密接触，滑动夹使 DNA 聚合酶能迅速地重新结合到同一引物-模板接头上，大大增加了 DNA 聚合酶的延伸能力。

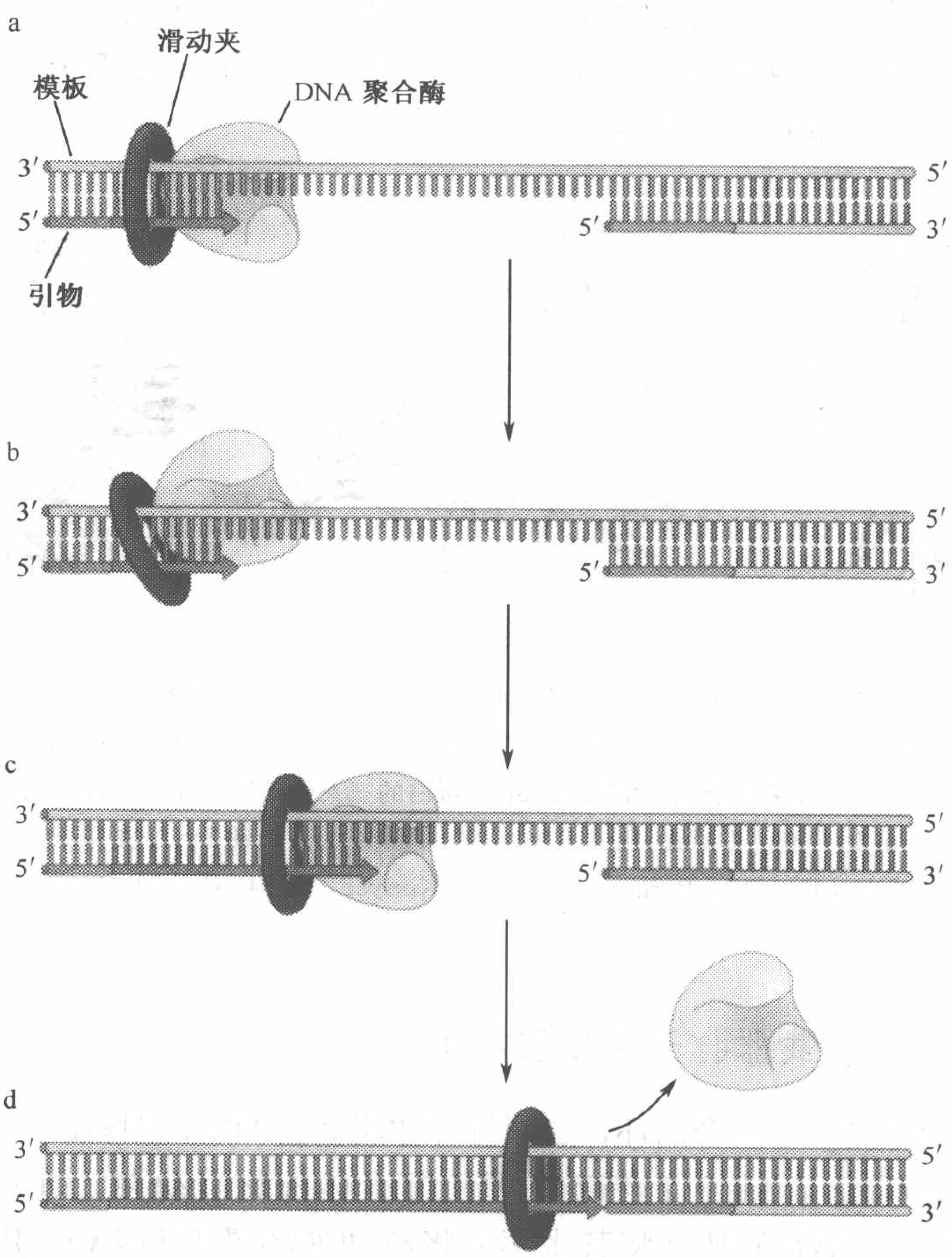


图 8-19　滑动 DNA 夹增强了相结合的 DNA 聚合酶的延伸能力。

当单链 DNA 模板完全复制后，DNA 聚合酶必须从此 DNA 和滑动夹上释放下来以在新的引物-模板接头处工作。此释放是通过依赖于环结合 DNA 的 DNA 聚合酶与滑动夹之间亲和力的变化实现的。与引物-模板接头结合的 DNA 聚合酶与夹子有很高的亲和力。但是，当 DNA 聚合酶到达单链 DNA 模板的末端时（如冈崎片段的末端），DNA 聚合酶构象发生变化，降低了其与滑动夹以及 DNA 之间的亲和力（图 8-19）。因此，当聚合酶完成一段 DNA 复制的时候，它即被滑动夹释放以使其可在新的引物-模板接头上工作。而夹子仍与 DNA 保持结合，并与在新合成 DNA 上作用的其他酶结合。

释放 DNA 聚合酶后，滑动夹并不立即从复制的 DNA 上脱落下来。相反，其他必须在 DNA 新合成处作用的蛋白质通过与夹子蛋白相互作用而执行其功能。如第 7 章所述，真核细胞中组装染色质的酶通过与真核滑动 DNA 夹（称为 PCNA）之间的相互作用被募集到 DNA 的复制部位。类似地，冈崎片段修复所涉及的真核蛋白也与滑动夹蛋白作用。不论何种情况，这些蛋白质均通过与滑动夹之间的相互作用在最需要它们的时候聚集到新的 DNA 合成部位。

滑动夹蛋白是病毒、细菌、酵母以及人类等生物中 DNA 复制体系中的保守部件。

与其功能保守相一致的是，源自这些不同生物的滑动夹的结构也是保守的（图 8-20）。不论何种情况，夹子均为同样的六边形对称并有相同的直径。尽管其整体结构非常类似，但形成夹子的亚基数目却不相同。

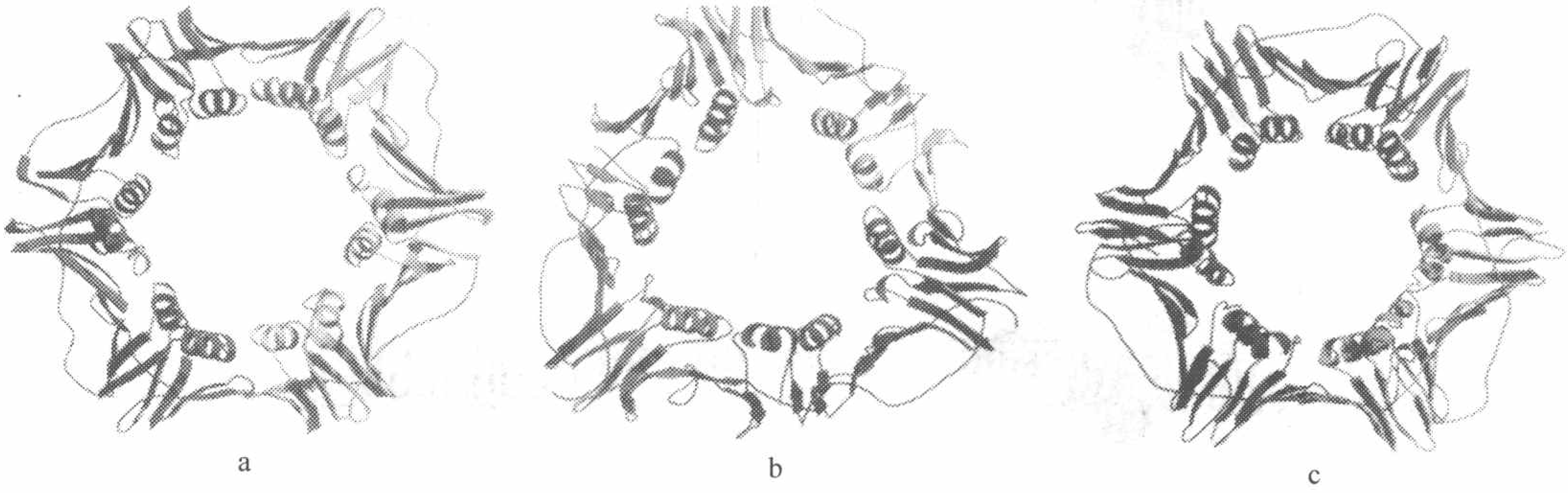


图 8-20 不同生物中分离得到的滑动 DNA 夹的三维结构。 所有生物中都发现有滑动 DNA 夹，它们有相似的结构。a. *E. coli* 的滑动 DNA 夹由两个拷贝的 β 蛋白构成。（Kong X. P.，Onrust R.，O'Donnell M.，and Kuriyan J. 1992. Cell 69:425.）。b. T4 噬菌体滑动 DNA 夹是 gp45 蛋白的三聚体（Moarefil.，Jeruzalmi D.，Turner J.，O'Donnell M.，and Kuriyan J. 2002. I Mol. Biol. 296:1215.）。c. 真核滑动 DNA 夹是 PCNA 蛋白的三聚体。（Krishna T. S.，Kong X. P.，Gary S.，Burgers P. M.，and Kuriyan J. 1994. Cell 79:1233.）镜像图用 BobScript、MolScript 和 Raster 3D 绘制。

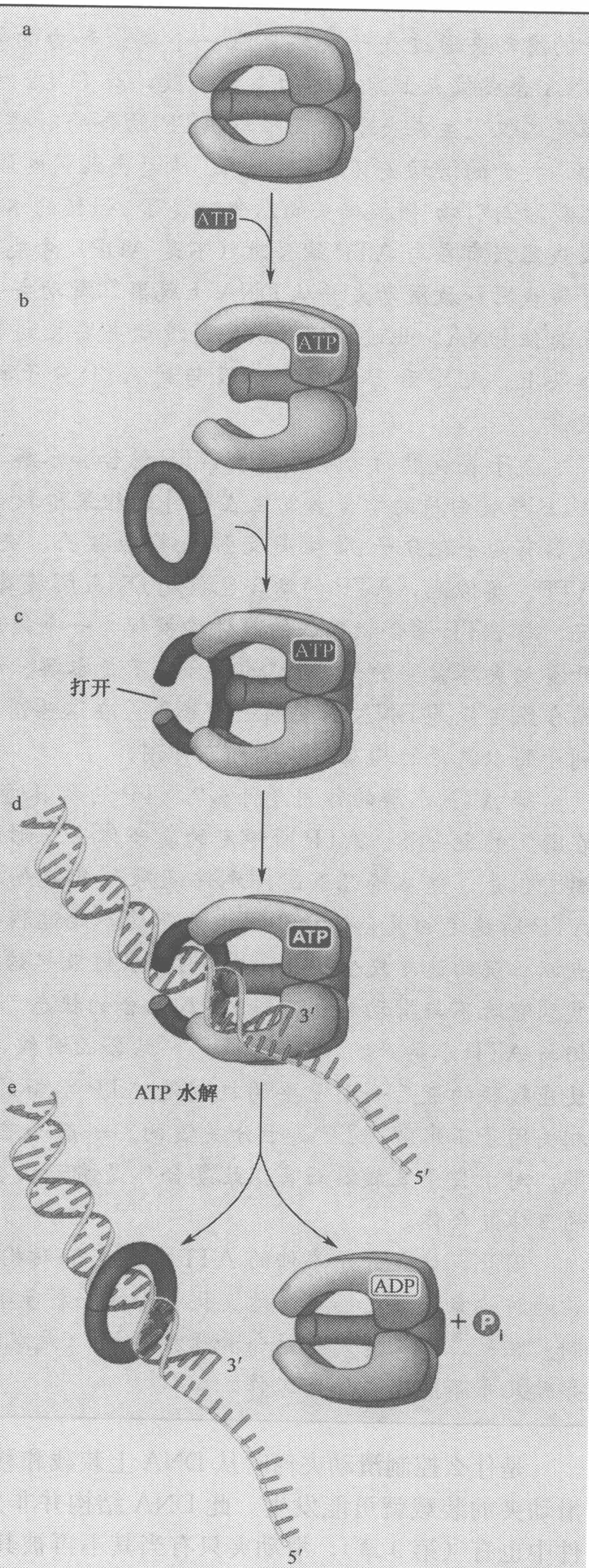
滑动夹通过滑动夹装载器打开并安置在 DNA 上

滑动夹在溶液中是一个闭合的环，必须打开才能套在 DNA 双螺旋上。一类称为滑动夹装载器（sliding clamp loader）的特殊的蛋白复合体催化滑动夹的打开并将其安放在 DNA 上。这些酶通过结合 ATP 并将其水解而将滑动夹环置于 DNA 的引物-模板接头上（框 8-4）。夹子装载器还能将滑动夹在不使用的时候将其从 DNA 上移除，像 DNA 解旋酶和拓扑异构酶一样，这些酶改变其靶物（滑动夹）的构象，但并不改变其化学组成。

框 8-4 ATP 控制蛋白的功能：滑动夹的装载

滑动 DNA 夹装载器由 5 个亚基组成，是一类含有相关 ATP 结合和水解位点的蛋白质家族成员，称为 AAA+蛋白。这种蛋白组合成 AAA+蛋白复合体，利用水解 ATP 的能量来改变目标蛋白和 DNA 的结构。滑动 DNA 夹装载器是由不同的 AAA+蛋白组成的，但是其他的 AAA+蛋白复合体一般是由一个 AAA+蛋白的多拷贝组成。尽管 AAA+蛋白含有相关 ATP 结合和水解位点，但是它们行使不同的功能。除了装载成 DNA 滑动夹，AAA+蛋白还解开 DNA，在 DNA 模板周围加载解旋酶（比如 ORC 和 DnaC，我们应该记得，DNA 复制起始时我们谈论过），展开蛋白质三级结构，拆分蛋白质复合物。实际上，因为它具有多个功能，我们称之为三磷酸腺苷酶——它和多种活性相联系。

ATP 的结合以及水解是如何与滑动夹的装载联系到一起的？当与 ATP 结合时，夹子装载器与滑动夹结合并通过拆开一个亚基-亚基界面使滑动夹打开（框 8-4 图 1）。打



框 8-4 图 1　滑动夹装载的 ATP 控制。 a. 滑动夹装载器是五亚基的蛋白质复合体，其活性是由 ATP 的结合和水解控制的。在 *E. coli* 中夹子装载器称为 γ-复合体，在真核细胞中称为复制因子 C（RF-C）。b. 为了催化滑动夹的打开，夹子装载器必须与 ATP 结合。c. 夹子装载器一旦与 ATP 结合，即与夹子结合并在一个亚基-亚基界面处将环打开。d. 形成的复合体可与 DNA 结合。DNA 的结合是由优先与引物-模板接头结合的夹子装载器介导的。与 DNA 的正确结合有两个结果：第一，打开后的滑动夹被定位，可使 DNA 处于夹子的孔洞之中；第二，DNA 的结合通过夹子装载器激发了 ATP 的水解。e. 因为只有结合 ATP 的夹子装载器才能与夹子以及 DNA 结合，所以结合 ADP 的夹子装载器很快就从夹子以及 DNA 上脱离下来，留下套在双链 DNA 引物-模板接头位置处的闭合的夹子。（来源：Based on O′Donnell M. et al. 2001. Clamp loader structure predicts the architectune of DNA polymerase Ⅲ holoenzyme and RFC. Current Biology 11: R942, Fig 5. © Elsevier.）

开的滑动夹通过夹子装载器上一个高亲和力的 DNA 结合位点被运送到 DNA 上。与 DNA 合成位点上需要滑动夹相一致，此 DNA 结合位点特异性地识别引物-模板接头，但是这仅发生在夹子装载器与 ATP 结合的时候。当夹子装载器与引物-模板接头结合时，打开的滑动夹环绕住 DNA。滑动夹装载的最后一步是由 ATP 水解所催化的。夹子装载器与引物-模板接头的结合激发了 ATP 的水解（通过夹子装载器进行）。因为夹子装载器只有在与 ATP 结合时（不是 ADP）才能与滑动夹及 DNA 结合，所以水解使夹子装载器释放滑动夹并从 DNA 上脱离。滑动夹一旦从夹子装载器上释放，即自动闭合而套住 DNA。此过程的净结果是滑动夹安装到 DNA 聚合酶的作用位点——引物-模板接头上。ADP 和 Pi 的释放以及与新 ATP 分子的结合使夹子装载器能够启动新一轮的装载。

夹子装载器的功能解释了 ATP 结合和水解与分子事件耦联的几个一般性质。通常 ATP 与蛋白质的结合多发生在事件的组装阶段：因子与靶分子的结合。例如，夹子装载器有两个靶分子-滑动夹及引物-模板接头。夹子装载器与这两个靶物的结合都需要 ATP。类似地，ATP 的结合也激发 DNA 解旋酶与单链 DNA 结合的能力。不论哪种情况，与 ATP 结合相耦联的事件均被视为一个循环中的一部分。对夹子装载器而言，打开滑动夹环需要的是 ATP 的结合而不是水解。对于 DNA 解旋酶来说，与单链 DNA 的结合很可能是 DNA 解旋的关键事件。在这些情况下，与 ATP 的结合稳定了酶的构象，利于酶以特定的构象与底物相互作用。

那 ATP 水解的作用是什么？ATP 的水解通常与事件中的解离阶段相关：释放结合在酶上的靶分子。ATP 所稳定的复合体一旦形成，就一定会解离。这可能是简单的分离。但是，更多情况下是将元件还原到其起始状态（如溶液中游离的滑动夹）。如果 ATP 所稳定的复合体结合得十分牢固，此过程将很缓慢。为了保证在正确的时间、地点以正确的速率发生解离，由 ATP 水解来起始解离。例如，ATP 水解使夹子装载器恢复到它既不与滑动夹也不与 DNA 结合的状态。这种初始状态的回复可能会发生在当酶仍与 ATP 水解产物（ADP 和 P_i）结合的时候，或是需要其分离的时候。ATP 水解与反应耦联的最后一个重要的机制是 ATP 水解的引发。在所需要的复合体组装好之前，相关因子不水解 ATP 是十分关键的。一般而言，特定复合体的形成可引发 ATP 的水解。对于夹子装载器而言，此复合体是滑动夹、夹子装载器以及引物-模板接头所形成的三体复合体。

因此，这些分子事件的 ATP 控制与酶对构象变化的时相控制最直接相关。酶通过在两种构象状态之间有序地变化和形成一种重要的中间物引发 ATP 水解以完成这种控制。相反，如果酶只是结合和释放 ATP（无水解），则反应将频繁回到起始状态，这将导致几乎不能完成任何工作。

是什么控制滑动夹何时从 DNA 上装载和移除呢？只要细胞中出现引物-模板接头，滑动夹的装载就可能发生。此 DNA 结构并非只在 DNA 复制时形成，在 DNA 修复事件中也有（第 9 章）。滑动夹只有当其不再被其他酶使用时才能从 DNA 上移除。滑动

夹装载器和 DNA 聚合酶不能同时与滑动夹作用，因为它们在滑动夹的同一面上有重叠的结合位点。因此，与 DNA 聚合酶结合的滑动夹不会从 DNA 上移除。类似地，核小体组装因子、冈崎片段修复蛋白以及其他 DNA 修复蛋白均与夹子装载器一样与滑动夹的同一区域作用。因此，滑动夹只能在与其相作用的所有酶都完成工作后才能从 DNA 上移除。

复制叉上的 DNA 合成

在复制叉上前导链和后随链同时进行合成，这对在 DNA 合成期内限制细胞内单链 DNA 的数量有重要意义。当 DNA 的单链区断裂时，在染色体上会产生完全的断裂，这比双链 DNA 上的一条单链 DNA 断裂修复起来要困难得多。而且，修复这种类型的损伤经常会导致 DNA 的突变（第 9 章）。因此，限制 DNA 处于此状态下的时间是十分重要的。为了协调 DNA 两条链的复制，在复制叉上有多种 DNA 聚合酶在行使功能。

E. coli 中，这些聚合酶的协同作用是通过它们连接成一个巨大的多蛋白复合体而实现的。此复合体被称为 DNA 聚合酶Ⅲ全酶（图 8-21）。全酶是核心酶和对其有增强功能的其他元件的多蛋白复合体的总称。DNA 聚合酶Ⅲ全酶含有 2 个拷贝的核心 DNA 聚合酶Ⅲ和 1 个拷贝的五蛋白 γ-复合体（*E. coli* 滑动夹装载器）。虽然 γ-复合体只有一个拷贝，但是它却与核心 DNA 聚合酶Ⅲ的 2 个拷贝结合，这对于全酶的形成是至关重要的（图 8-21）。与两个拷贝的 DNA 聚合酶相结合的 γ-复合体的相同亚基同时与 DNA 解旋酶有较微弱的相互作用（下节）。

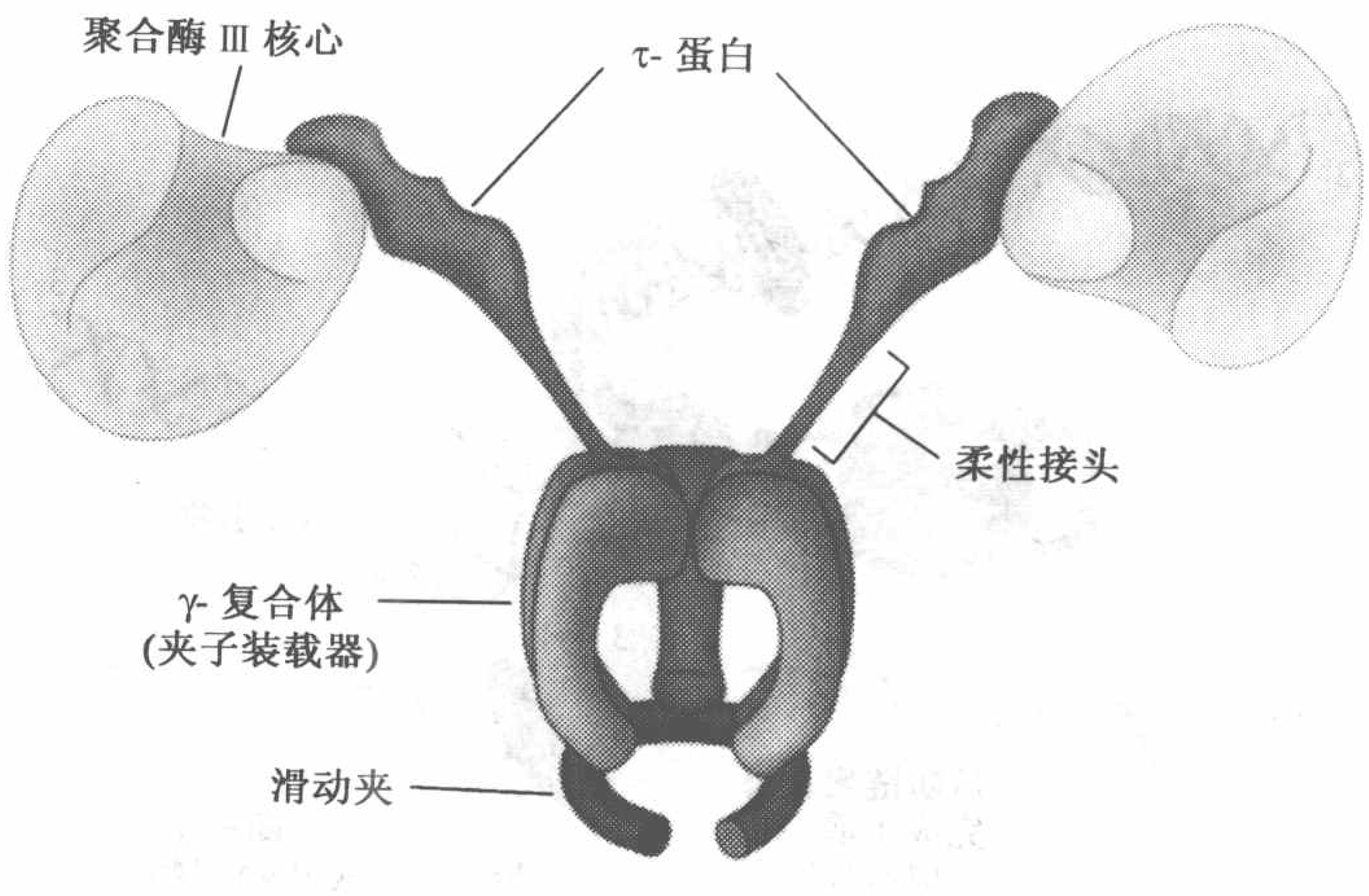
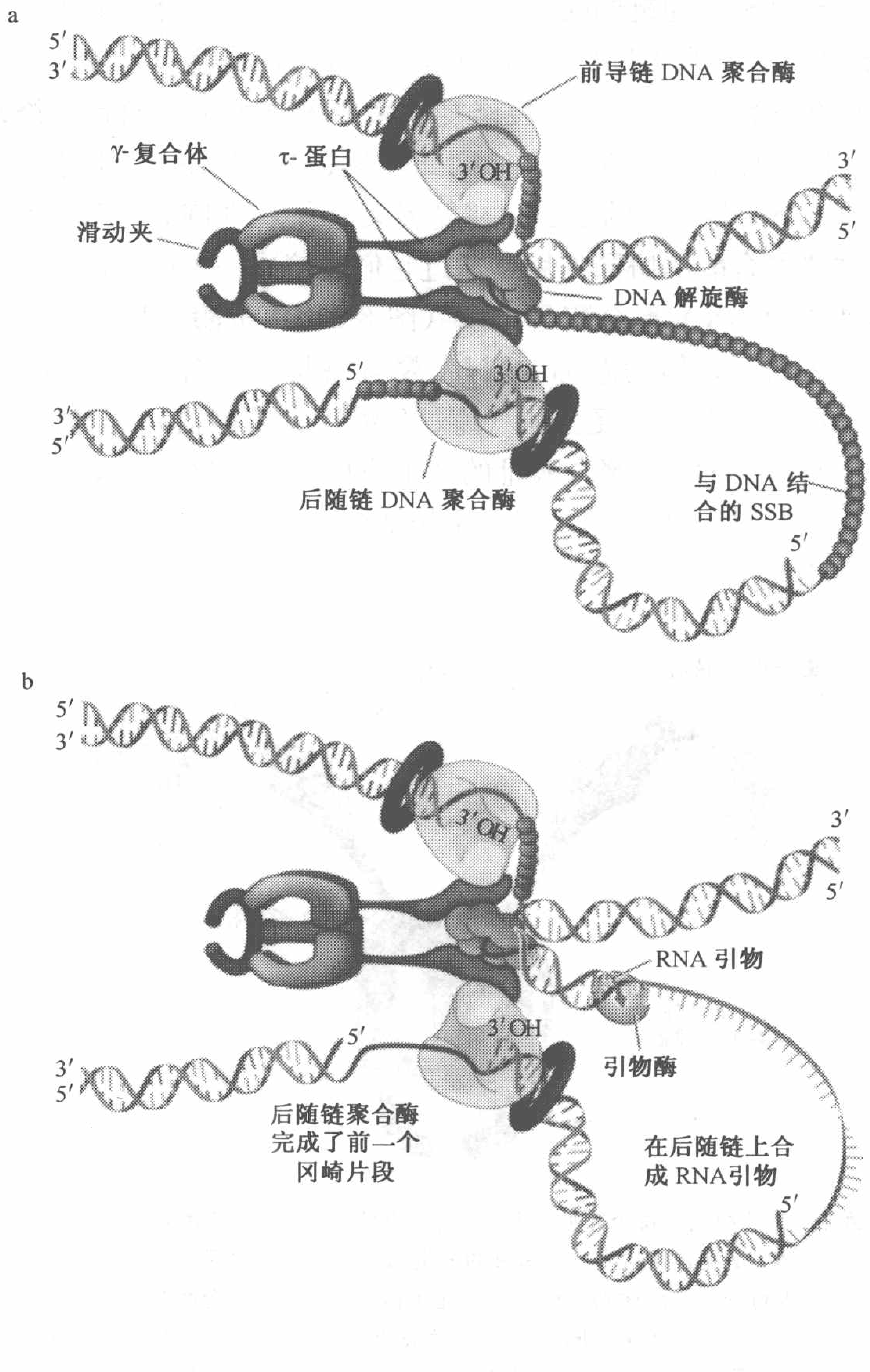


图 8-21　DNA 聚合酶Ⅲ全酶的组成。 每个拷贝的 DNA 聚合酶Ⅲ全酶中有 3 个酶：2 个拷贝的 DNA 聚合酶Ⅲ核心酶和 1 个拷贝的 γ-复合体。γ-复合体含有 2 个拷贝的 τ-蛋白，每个 τ-蛋白都有一个与一个 DNA 聚合酶Ⅲ核心酶相作用的域。对 τ-蛋白进行的氨基酸序列分析表明，此蛋白质的 DNA 聚合酶Ⅲ结合区域与夹子装载所涉及的蛋白质部分之间由一段延伸的柔性接头所间隔。这个接头据认为可以使两个聚合酶以相对独立的方式运动，这种方式对于使一个聚合酶复制前导链而另一个聚合酶复制后随链来说是必需的。（来源：Based on O'Donnell M. et al. 2001. Clmap loader structure predicts the architecture of DNA polymerase Ⅲ holoenzyme and RFC. Current Biology 11:R943, Fig. 6. ©Elsevier.）

当DNA同时在前导链和后随链模板上合成的时候，两个DNA聚合酶如何在复制叉处保持连接？一个解释此耦联效应的模型认为复制机器利用了DNA的柔性（图 8-22)。当解旋酶在复制叉处解开DNA螺旋时，前导链暴露出，并与前导链上的DNA聚合酶作用，合成一条连续的DNA互补链。相反，后随链模板并不是立即与DNA聚合酶作用，而是以单链DNA的形式伸出，并被SSB迅速结合。引物酶与DNA解旋酶相互作用从而被激活，在后随链模板上不连续地合成一条新的RNA引物。产生的RNA：DNA杂交体被滑动夹子装载器识别为引物-模板接头．滑动夹子在此处组装。



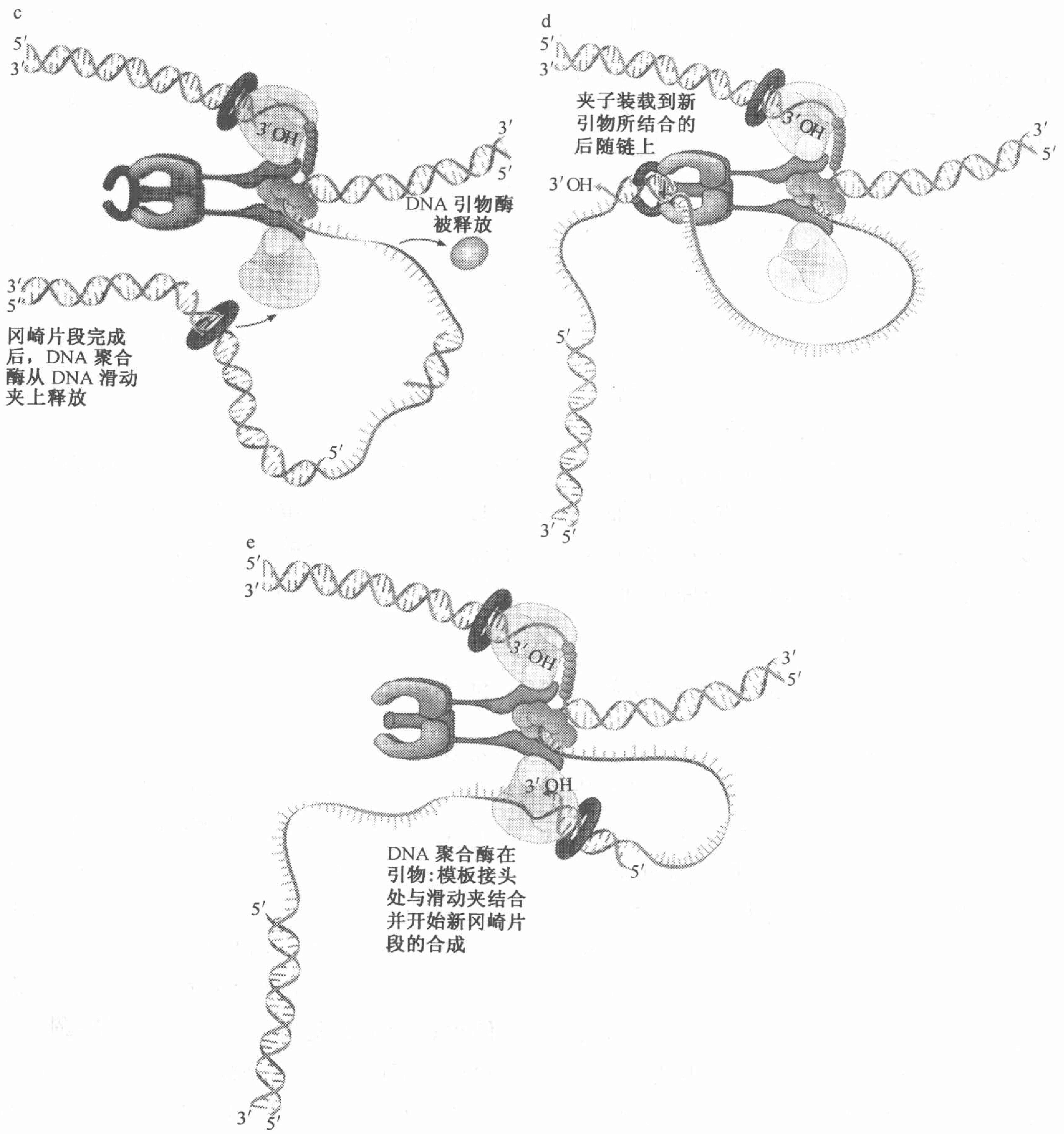


图 8-22 ***E. coli*** **复制叉上用于协调两种 DNA 聚合酶复制的"长号模型"。** a. *E. coli* DNA 复制叉上的 DNA 解旋酶在后随链模板上沿着5′→3′方向运动。DNA 聚合酶Ⅲ全酶通过 τ-亚基与 DNA 解旋酶相互作用，τ-亚基与两个 DNA 聚合酶都结合。一个 DNA 聚合酶Ⅲ核心酶复制前导链，另一个 DNA 聚合酶Ⅲ核心酶复制后随链。SSB 覆盖在 DNA 的单链区域（为了简便起见，后随链上的 SSB 只在 a 部分中显示）。b. DNA 引物酶周期性地与 DNA 解旋酶结合并在后随链模板上合成新的引物。c. 当后随链 DNA 聚合酶完成一个冈崎片段之后即从滑动夹和 DNA 上脱落。d. 随后，刚刚被引物结合的后随链 DNA 成为夹子装载器的靶子。夹子装载器在由新合成的 RNA 引物生成的引物-模板接头位置上组装成新的滑动夹。e. 结合有滑动夹的引物-模板接头与后随链 DNA 聚合酶结合，起始下一个冈崎片段 DNA 的合成。虽然这个描述着重于较复杂的后随链在合成过程中的作用，但在整个过程中，前导链上新单链 DNA 模板也已经产生并由前导链上的 DNA 聚合酶Ⅲ进行快速地复制。

当后随链上的 DNA 聚合酶完成前面的冈崎片段时，即从模板上释放。因为该聚合酶与前导链上的 DNA 聚合酶和滑动夹子装载器保持着联系，所以在加入滑动夹子之后该 DNA 聚合酶立即与 RNA 引物-模板接头在理想的位置结合。由于后随链聚合酶合成一条冈崎片段的时间比 RNA 引物的合成和滑动夹子组装的时间短，后随链 DNA 聚合酶准备并等待滑动 DNA 夹子装载器（γ 复合体）释放引物-模板接头。通过和引物-模板接头的结合，后随链复制每形成一个新环并启动下一轮的冈崎片段的合成。这个模型被称作“长号模型”，以比喻在后随链模板上 DNA 聚合酶和 DNA 解旋酶之间形成的单链 DNA 环状结构大小的变化。

真核细胞中 DNA 的复制也需要多种 DNA 聚合酶。每个复制叉上有 3 种不同的 DNA 聚合酶：DNA 聚合酶 α/引物酶、DNA 聚合酶以及 DNA 聚合酶 ε（图 8-17）。DNA 聚合酶 α/引物酶启动新链合成，DNA 聚合酶 δ 利 DNA 聚合酶 ε 进行链的延伸。近来的研究提供了强有力的证据表明 DNA 聚合酶 ε 合成前导链，DNA 聚合酶 δ 合成后随链。已知许多其他蛋白质为真核细胞复制叉的一部分。这些蛋白质的作用目前了解甚少。然而，他们很大可能协调三种 DNA 聚合酶，并耦联聚合酶和真核细胞 DNA 解旋酶（Mcm2-7）相互作用。与原核细胞所不同的是，真核细胞的滑动夹子装载器，RFC，不行使这样的功能。

复制叉蛋白之间的相互作用形成 *E. coli* 的复制体

DNA 聚合酶Ⅲ全酶组分之间的联系并不是发生在细菌复制叉各元件中唯一的相互作用。除了 DNA 聚合酶Ⅲ全酶组分之间的相互作用以外，还有几个蛋白质-蛋白质间的相互作用用于加快复制叉的进程。其中最重要的是 DNA 解旋酶（六聚体 dnaB 蛋白，表 8-1）与 DNA 聚合酶Ⅲ全酶（图 8-23）之间的相互作用。这个由全酶的夹子装载器介导的相互作用将解旋酶和 DNA 聚合酶Ⅲ全酶连到一起。此结合使解旋酶运动的速度增加 10 倍，从而激发了解旋酶的活性。如果 DNA 解旋酶与 DNA 聚合酶分离，其速度将减慢（图 8-23）。解旋酶活性与 DNA 聚合酶Ⅲ全酶耦联，防止了解旋酶从 DNA 聚合酶Ⅲ全酶上“逃脱出去”，使得这两个关键的复制叉酶得以协调。

第二个重要的蛋白质间相互作用发生在 DNA 解旋酶与引物酶之间。与 *E. coli* 复制叉中的多数蛋白质不同，引物酶与复制叉的结合并不紧密。相反，在约 1 秒一次的间隔中，引物酶与解旋酶以及 SSB 覆盖的单链 DNA 结合并合成新的 RNA 引物。虽然 DNA 解旋酶与引物酶之间的作用相对较弱，但是其相互作用大大激发了引物酶的功能（提高约 1000 倍）。RNA 引物合成后，引物酶从 DNA 解旋酶上脱落并进入溶液中。

在 *E. coli* 中引物酶与 DNA 解旋酶之间相对较弱的相互作用对于调节冈崎片段长度很重要。紧密的结合可导致后随链上的引物合成更频繁，这样冈崎片段就会较短。较弱的作用则产生较长的冈崎片段。

复制叉上发挥作用的全部蛋白质的总和称为复制体（replisome）。这些蛋白质在一起形成一个用于 DNA 合成的调度有序的工厂，此工厂含有多个相互关联的机器。这些机器都可单独地行使重要的特殊功能。当它们在一起的时候，其活动通过相互之间的作

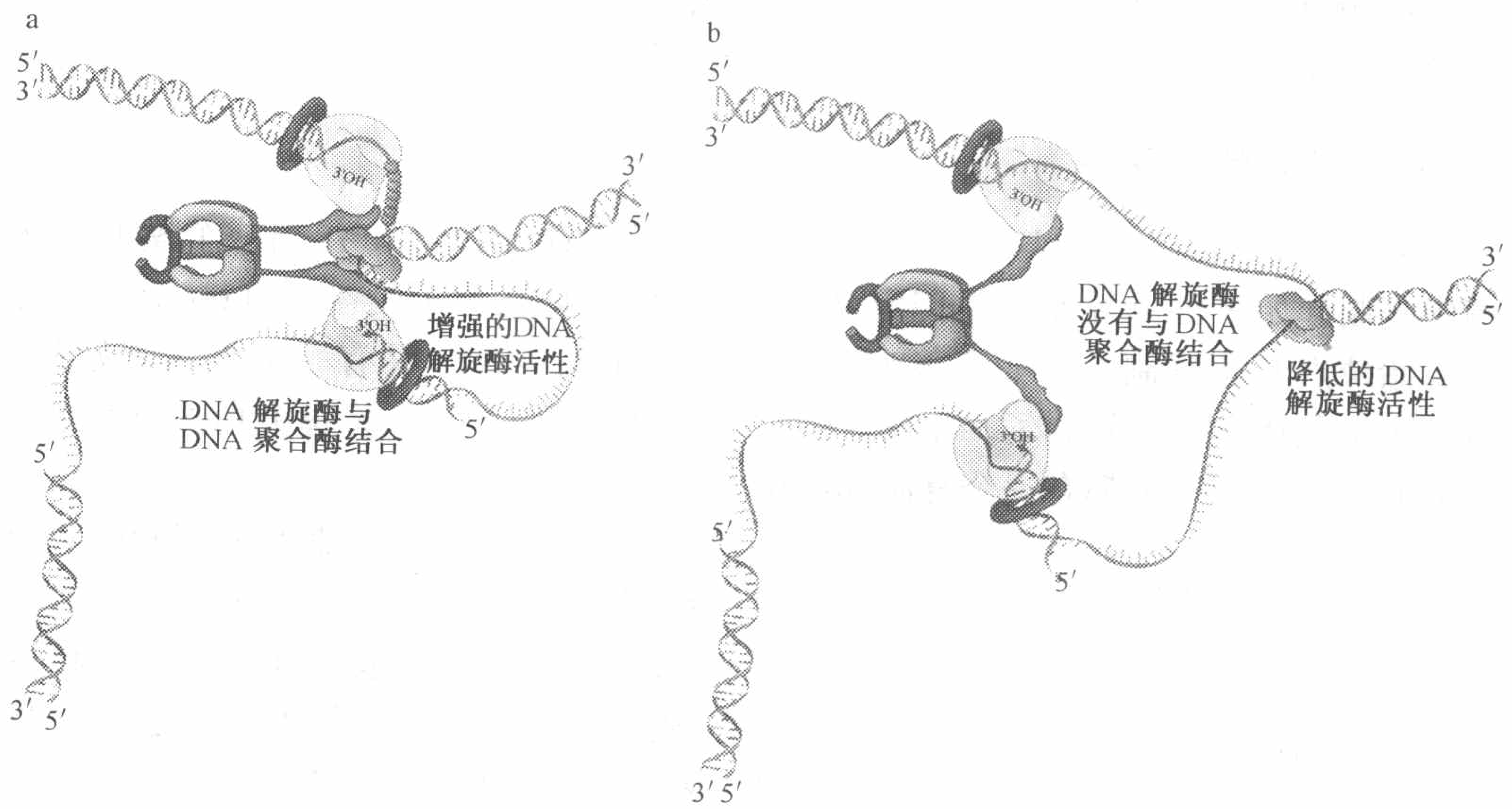


图 8-23 DNA 解旋酶与 DNA 聚合酶Ⅲ全酶的结合加快 DNA 链分离的速度。夹子装载器的 τ-亚基在复制叉上与 DNA 解旋酶和 DNA 聚合酶都发生作用。a. 当此相互作用形成后，DNA 解旋酶以与 DNA 聚合酶复制 DNA 大致同样的速度解开 DNA。b. 如果 DNA 解旋酶没有与 DNA 聚合酶Ⅲ全酶作用，则 DNA 解开的速度变为原来的 1/10。在此情况下，DNA 聚合酶的复制快于 DNA 解旋酶对未复制 DNA 链的拆分。这使 DNA 聚合酶Ⅲ全酶能赶上 DNA 解旋酶并重新形成一个完整的复制体。

用进行协调。这些相互作用在 *E. coli* 细胞中已所知甚详，但是对噬菌体和真核 DNA 复制机器的研究表明，这些生物中的 DNA 复制亦涉及类似的多机器间的协调作用。事实上，在 *E. coli* 复制中涉及的蛋白质在其他生物中有明确的同源蛋白质。噬菌体、原核生物和真核生物 DNA 复制中执行类似功能的因子的名称见表 8-1。

为了更好地体会复制 DNA 的酶令人惊叹的能力，设想一个 DNA 碱基有你的课本那么大。这种情况下双链 DNA 将有约 1m 的直径，*E. coli* 基因组将是一个约 500 英里（800 km）周长的大环。更重要的是，复制体有联邦快递的运货车那么大，它将以超过 375 英里/h（600 km/h）的速度运动！对两个这样的机器来说，*E. coli* 基因组的复制为历时 40min、250 英里（400km）的路程，每个机器在它的航迹上留下两条 1m 粗的 DNA 线缆。令人叫绝的是，复制机器在路程当中平均仅产生一个错误。

DNA 复制的起始

特定的基因组 DNA 序列指导 DNA 复制的起始

复制叉最初的形成需要 DNA 双链的分离提供单链 DNA 用于 DNA 解旋酶结合和 RNA 引物及新 DNA 的合成所必需的模板。虽然 DNA 链的分离（也称为 DNA 解旋）在染色体末端最容易进行，但是 DNA 的合成通常是在中间区域起始的。事实上，对于环状染色体来说，无染色体末端使内部 DNA 解旋对复制起始很重要。

DNA 解旋和复制起始发生的特殊位点称作复制起始位点（origin of replication）。依不同的生物而不同，每个染色体上可有少至一个多至数千个的起始位点。

复制起始的复制子模型

1963 年，Francois Jacob、Sydney Brener 以及 Jacques Cuzin 提出一个模型来解释细菌中复制起始事件的控制。他们定义所有的 DNA 均从称为复制子（replicon）的特定的起始位点开始复制。例如，因为 *E. coli* 中的单条染色体只有一个复制起始位点，所以整个染色体是一个复制子。相反，多个复制起始位点的存在将每个真核染色体分成多个复制子——每一个都有一个复制起始位点。

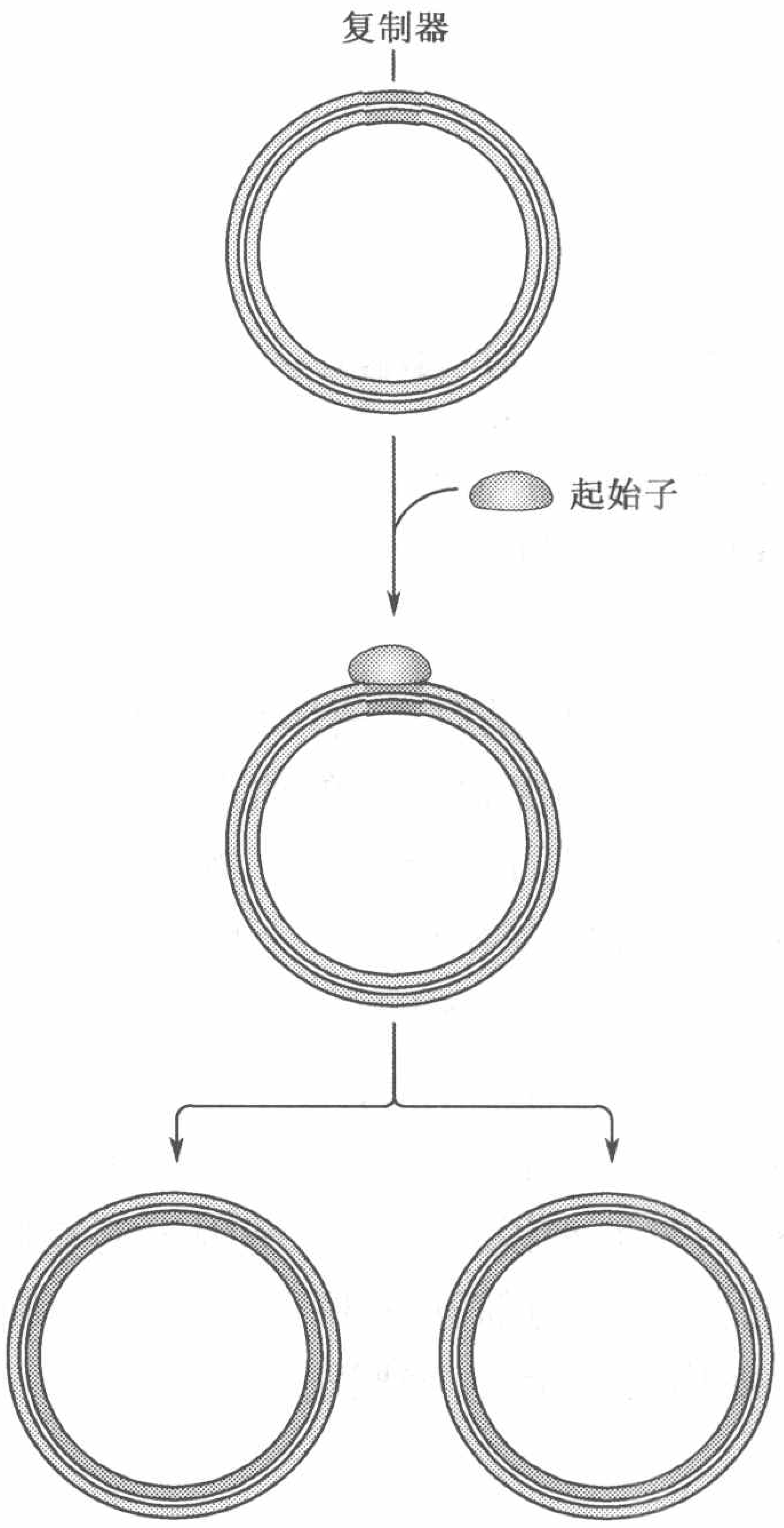


图 8-24 复制子模型。起始子与复制器的结合引发了复制的起始和相连 DNA 的复制。

复制子模型提出了控制复制起始的两个元件：复制器和起始子（图 8-24）。复制器（replicator）是指足够指导 DNA 复制起始的整套顺式作用的 DNA 序列。与之相对照的是，复制起始位点是在 DNA 上发生 DNA 解旋以及 DNA 合成起始部位的位点。尽管复制起始位点始终是复制器的一部分，但是有时（尤其是在真核细胞中）复制起始位点仅仅是指导复制起始（复制器）的 DNA 序列的一部分。如我们将在第 12 章所见到的，在转录启动子与转录起始位点之间也有同样的区别。

复制子模型的第二个元件是起始子（initiator）蛋白。此蛋白质特异性地识别复制器中的一个 DNA 元件并激活复制的起始(图 8-24)。起始子蛋白质已经在多种生物中得到确认，包括细菌、病毒以及真核细胞。虽然这些蛋白质用不同的 DNA 结合基序识别 DNA，但是他们都选择的位点将成为复制起点。有趣的是，所有已知起始子蛋白质均由 ATP 结合和水解蛋白调节，并且有共同的 ATP 结合基序。该基序和 DNA 滑动夹装载器不同但相关联。

如下面我们所看到的，起始子蛋白是复制起始中涉及的唯一的序列特异性 DNA 结合蛋白，复制起始所需的其他蛋白质均不与 DNA 特异性地结合。相反，这些蛋白质通过相互作用以及对特殊 DNA 结构（如单链 DNA 或引物-模板接头）的亲和性而聚集到起始器上。

复制器序列包括起始子结合位点和容易解旋的DNA

复制器的 DNA 序列有 2 个共同的特性（图 8-25）。第一，它们含有起始子蛋白质结合位点，此位点是组装复制起始机器的核心；第二，它们含有一段富含 AT 的 DNA，此段 DNA 容易解旋但并不自发进行。在复制器上，DNA 的解旋是由复制起始蛋白控制的，多数生物中这些蛋白质被严格调控。

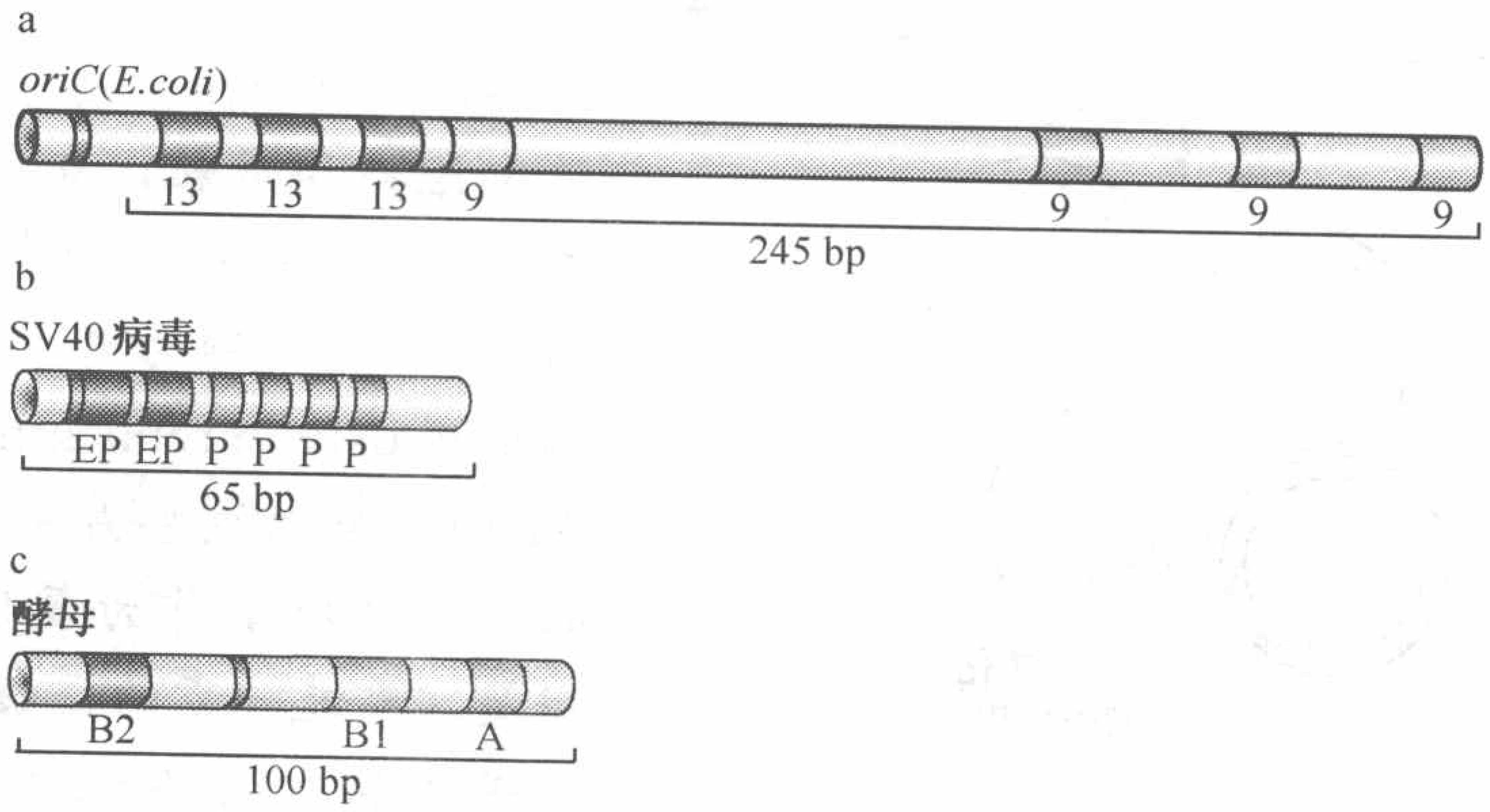


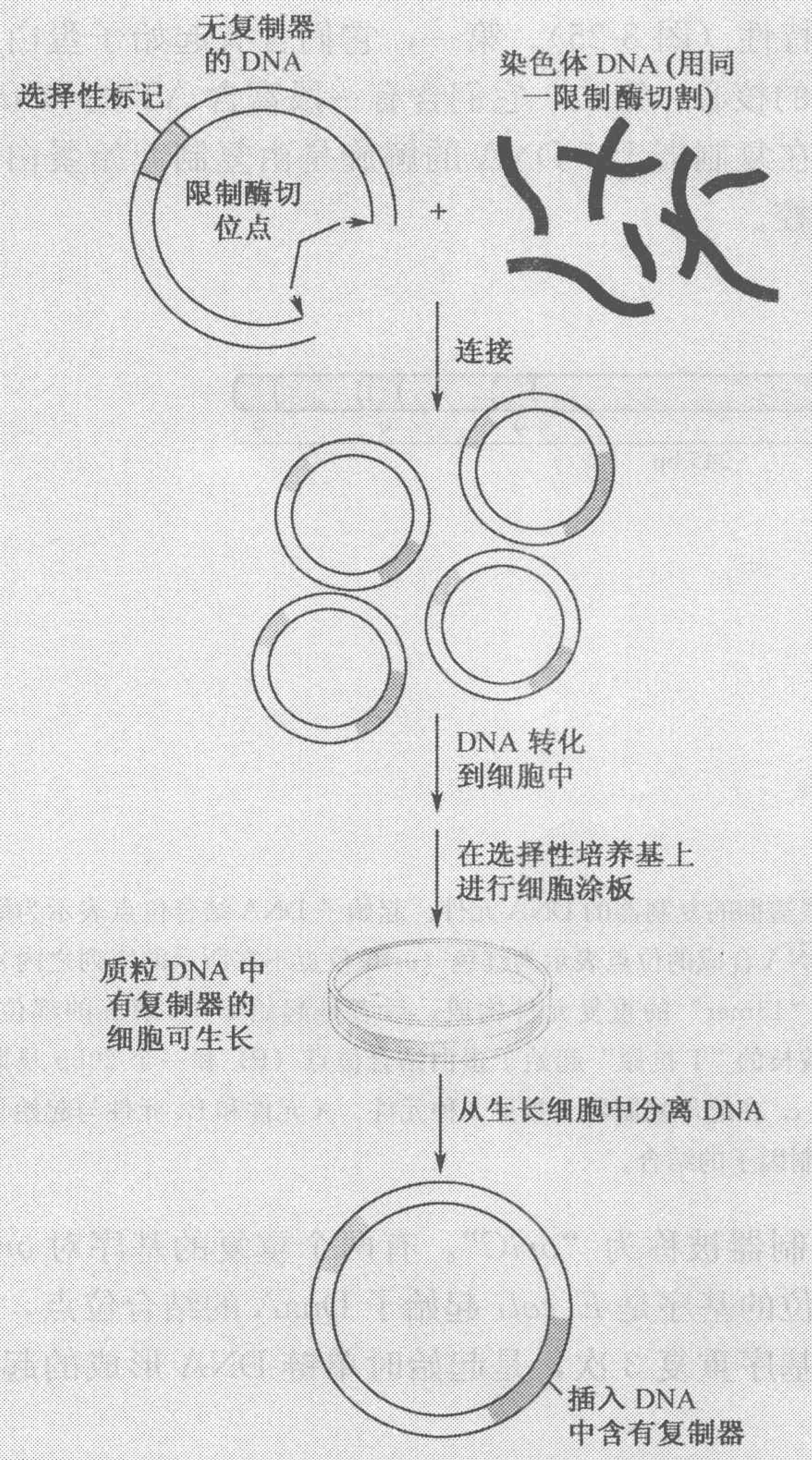
图 8-25 复制器的结构。图示为组成 3 个研究得很透彻的复制器的 DNA 元件。起始子 DNA 结合位点表示为绿色，促进 DNA 解旋的元件表示为蓝色，第一段 DNA 合成的位点表示为红色（*oriC* 位点不在所示的序列之内）。a. *oriC* 由 5 个“9mer”的 DnaA 结合位点和 3 个“13mer”的重复元件构成，后者是起始 DNA 解旋的部位。b. 真核病毒 SV40 的复制起始位点由 4 个 5 核苷酸长的“T 抗原”起始子蛋白结合位点（P）和一个 20bp 早期回文序列（EP）构成，后者是 DNA 解旋的部位。c. 在酵母复制器上常见 3 种元件。A 元件和 B1 元件与起始子 ORC 结合，B2 元件促进 DNA 的解旋和与其他复制因子的结合。

E. coli 染色体复制所需要的单个复制器被称为“*oriC*”。有两个重复的基序对 *oriC* 功能至关重要（图 8-25a）。9 核苷酸单位的基序是 *E. coli* 起始子 DnaA 的结合位点，其在 *oriC* 上重复 5 次。13 核苷酸单位的基序重复 3 次，是起始时单链 DNA 形成的起始位点。

虽然很多真核生物病毒和单细胞真核酿酒酵母复制器的特殊序列不同，但其整体结构却很相似（图 8-25b，c）。定义复制起始位点的方法见框 8-5。

对多细胞真核生物的复制器了解还不是很多。由于缺少在单细胞真核生物和细菌中用于复制起始位点鉴别的可稳定繁殖的环形小 DNA，对多细胞真核生物复制器的鉴别和特征的了解变得困难重重（框 8-5）。在少数几例得到鉴别的复制器中，发现它们比酿酒酵母和细菌染色体中的复制器大得多，长度通常超过 1000bp。与那些较小的复制器不同，很难分离到使这些复制器功能丧失的突变，这可能是因为在这些序列中重要的元件是冗余的。

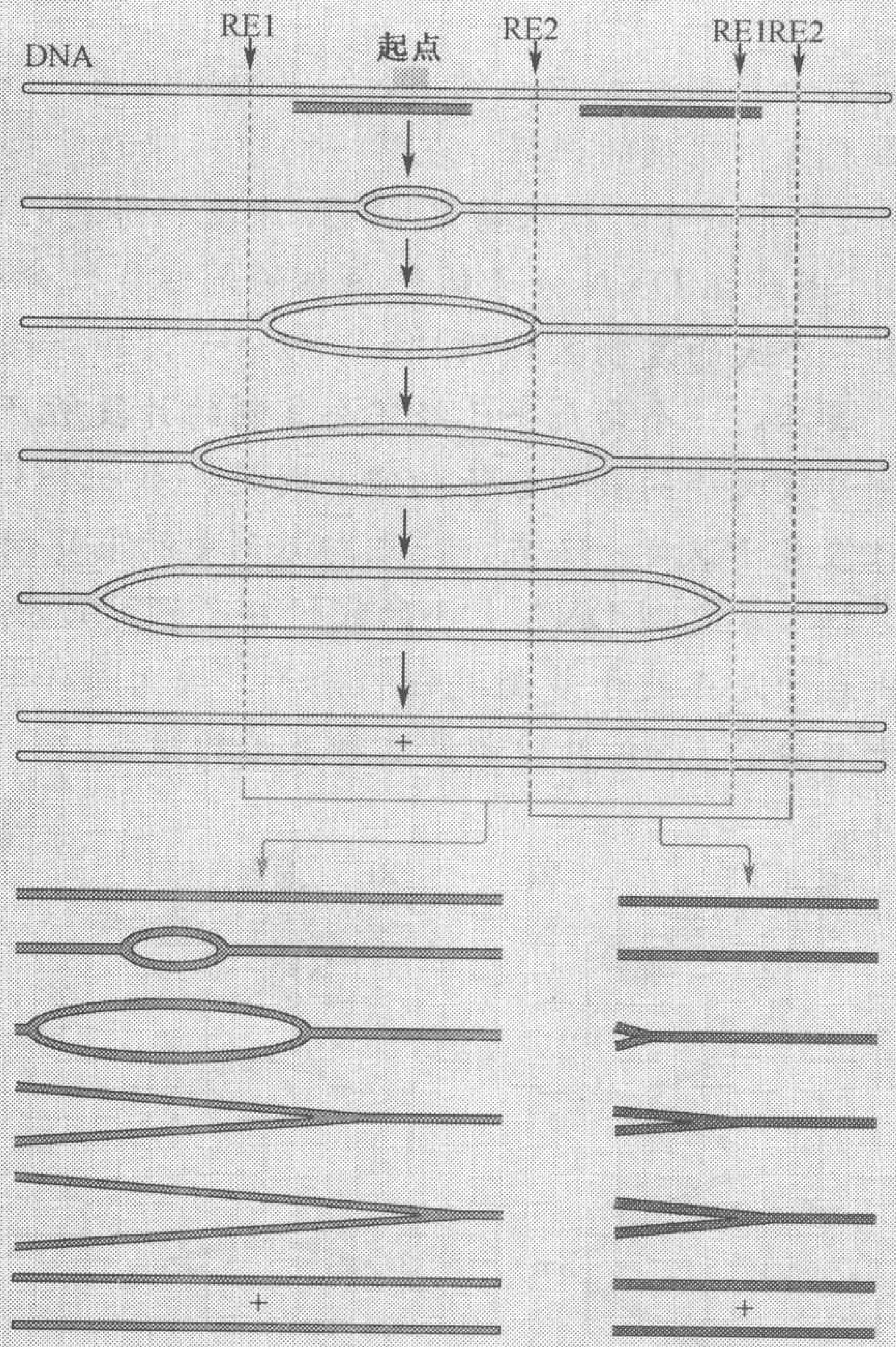
框 8-5　复制起始位点和复制器的鉴别



框 8-5 图 1　复制器的遗传学鉴别。用限制酶切割含有选择性标记的质粒（小的环形 DNA 分子），使质粒中正常的复制器被切除，产生一个不含复制器的 DNA 片段。为了从特定物种中分离到复制器，使用同一种限制酶切割来自此物种的 DNA 并将其连接到被切割的质粒中，重新生成环形的质粒，每个质粒中都含有一个来自受试物种的片段。此 DNA 然后转化到宿主中并用质粒上的选择性标记对重组的质粒进行筛选（如果标记为抗生素抗性，细胞将在有抗生素情况下生长）。生长的细胞中含有质粒及其选择性标记，表明质粒能够在细胞中复制而且质粒一定含有复制器。从宿主细胞中分离质粒并对插入 DNA 进行测序，可鉴别含有复制器的片段的序列。进一步对插入 DNA 进行诱变（如删除插入 DNA 的特殊区域），并重复上述实验可对复制器进行精细地鉴定。

复制器序列通常用遗传学实验进行鉴别。例如，首批酵母复制器就是用 DNA 转化实验鉴别的（框 8-5 图 1）。在这些研究中，研究人员将基因组 DNA 片段随机地克隆到缺少复制器但含有一个宿主细胞中缺少的选择性标记的质粒中。对于转化后可在细胞中维持的质粒，克隆的 DNA 片段中一定含有一个酵母复制器。被鉴别的 DNA 片段称为自主复制序列（autonomously replicating sequence，ARS）。虽然这些序列是在环形质粒的人工环境下作为复制器而作用的，但是需要进一步的证据来证明这些序列在其原本的染色体位置处也是复制器。

为了证实这些 DNA 序列在染色体中是作为复制器而起作用的，有必要发展用于鉴定复制起始位点在细胞中定位的方法。一个方法是利用在复制起始过程中形成的 DNA 复制中间体不寻常的结构。与完全复制的或完全未复制的 DNA 不同，正处于复制过程中的 DNA 不是线性的。例如，不含有复制起始位点的 DNA 片段（由限制酶对 DNA 切割产生）在其复制的时候具有各种“Y”形状的构象（框 8-5 图 2，蓝色的 DNA 片段）。类似地，含有复制起始位点的 DNA 片段复制起始后立即呈“水泡”形状。最后，如果复制起始位点不对称地位于 DNA 片段内，则 DNA 开始呈水泡形状然后转变成“Y”形状（框 8-5 图 2，红色的 DNA 片段）。利用双向琼脂糖凝胶电泳可将这些特殊形状的 DNA 与大多数线性 DNA 区分开来（框 8-5 图 3）。

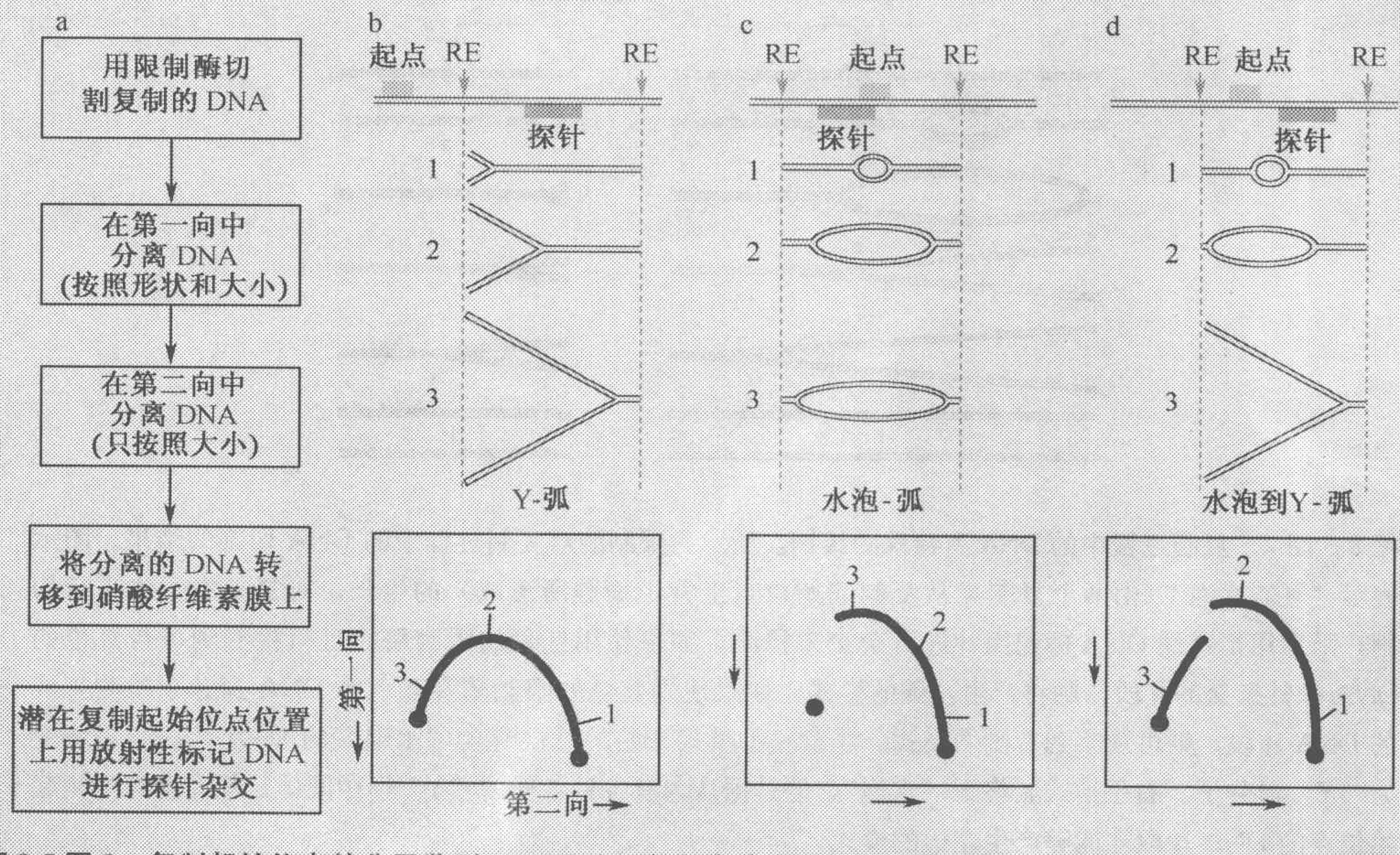


框 8-5 图 2 复制过程中的 DNA 有特殊的结构。所示为限制酶对复制过程中的 DNA 切割的结果。图示表示"复制水泡"(由两个复制叉从复制起始位点上背道而驰所形成)的生长。对复制中间体进行切割，随后用指示性 DNA 标记探针进行杂交来检测。如果使用红色的限制酶而且只检测到可与红色的 DNA 探针杂交的片段，则将产生左侧的模式。如果使用蓝色的限制酶和蓝色的 DNA 探针来检测生成的 DNA 片段，则可观察到右侧的模式。注意左手模式开始的时候有含"水泡"的 DNA 片段，最终变为"Y"形分子。右手模式始终无"水泡"，但是据认为有各种各样的"Y"形中间体。只有含有复制起始位点的 DNA 片段才能够产生左边的模式。

为了鉴别在复制过程中的 DNA，首先用限制酶将正在分裂细胞的 DNA 切断并用双向琼脂糖凝胶电泳进行分离。在第一向中，DNA 按照大小和形状进行分离，在第二向中，DNA 主要按照大小分离。这是通过在每一向上使用不同的凝胶密度和电泳速率达到的。为了按照大小和形状进行分离，琼脂糖凝胶的孔径要小（高琼脂糖密度），而且电泳速度要快。相反，如果主要是按照大小进行分离，则琼脂糖凝胶的孔径要大（低琼脂糖密度）一些并且电泳的速度要慢一些。当电泳完成后，将 DNA 分子转移到硝酸纤维素膜上并用 Southern 印迹检测（第 21 章）。对限制酶和 DNA 探针的选择能极大地影响分析的效果。通常，这个方法需要研究人员对复制起始位点可能的定位有充分认识。近来，出现了利用 DNA 芯片鉴定起始位点的新方法，这种方法不需要任何对起始

位点位置的光验认识。

双向电泳是如何鉴别含复制起始位点的 DNA 中间体的呢？DNA 迁移的特殊模式能为复制起始位点的鉴别提供明确的证据。在第一向中特殊的结构迁移最慢。例如，一个含有 3 个等长臂的“Y”形分子，在从特定 DNA 片段中得到的所有这样的分子中迁移最慢（框 8-5 图 3b），因此在 DNA 分子弧形顶端的是非线性分子。相反，一个带有两个非常短的复制臂和一个大的复制区域的“Y”形分子，其迁移与同一 DNA 片段的未复制版本非常相似。最后，一个由几乎已经完全复制的片段所产生的“Y”形分子在形状上与 2 倍于未复制片段大小的线形分子相似。因此，当一个 DNA 分子被单个复制叉复制时，它迁移的位置异于这样一个点，此点与弧形中的未复制片段紧邻，而且此弧形最终到达的位置是 2 倍于未复制 DNA 大小的线性分子预计迁移到的位置。此形状被称为Y-弧（Y-arc），代表的是正处于复制过程的分子。因为每一轮复制中所有的 DNA 分子都被复制，所以大部分的 DNA 片段都是这种类型的模式。



框 8-5 图 3 复制起始位点的分子鉴别。 a. 通过双向电泳分离 DNA，可以将处于复制过程中的 DNA 与完全复制的或完全未复制的 DNA 分离开来。总 DNA 由分裂期细胞（即正在复制）提取得到。DNA 首先按照大小和形状进行分离（使用高电压电泳和相对小的孔径）。然后电场旋转 90°，DNA 再主要按照大小进行分离（在大孔径琼脂糖上用低电压电泳）。用 Southern 分析来检测插入的 DNA。图示为 3 种不同的模式。最大的复制水泡在第一向上迁移最慢（c），带有几乎等长臂的“Y”形分子迁移次慢（b）。因为“Y-弧”和“水泡-弧”模式很难区分，所以“水泡到 Y-弧”（模式 d）认为是对起始位点的最好显示。

含有复制起始位点的分子形成水泡形状的复制中间物，此中间物在第一向中迁移得甚至比“Y”形分子还慢。泡越大，迁移越有异于线性 DNA（框 8-5 图 3c）。不幸的是，对泡形复制中间物的片段所产生的弧（称为水泡弧）与“Y”形中间物产生的弧很难分辨（框 8-5 图 3b 和框 8-5 图 3c）。如果起始位点不对称地位于 DNA 片段内，则此

困难可得到解决。在这种情况下，中间物开始是泡状，但是，当最靠近片段末端的复制叉完成复制的时候，泡状的中间物将变成“Y”形。这个所谓的泡形到“Y”形转变(bubble-to-Y transition)造成弧的间断，很容易被检测到，可作为复制起始位点的明确标志（框 8-5 图 3d)。因此，理想化的状态是选择在待检的复制起始位点两侧不对称的位置上进行切割的限制酶。

结合和解旋：起始子蛋白对复制起始位点的选择和激活

起始子蛋白在复制起始过程中一般执行至少两种不同的功能（图 8-26)。第一，这些蛋白质与复制器中的特异 DNA 序列结合；第二，起始蛋白质与复制起始过程中另外的必需因子相互作用，将其聚集到复制器。不是所有的起始子蛋白质行使第三种功能：他们扭曲或解旋其结合位点附近的 DNA 区域以利于打开双链 DNA。

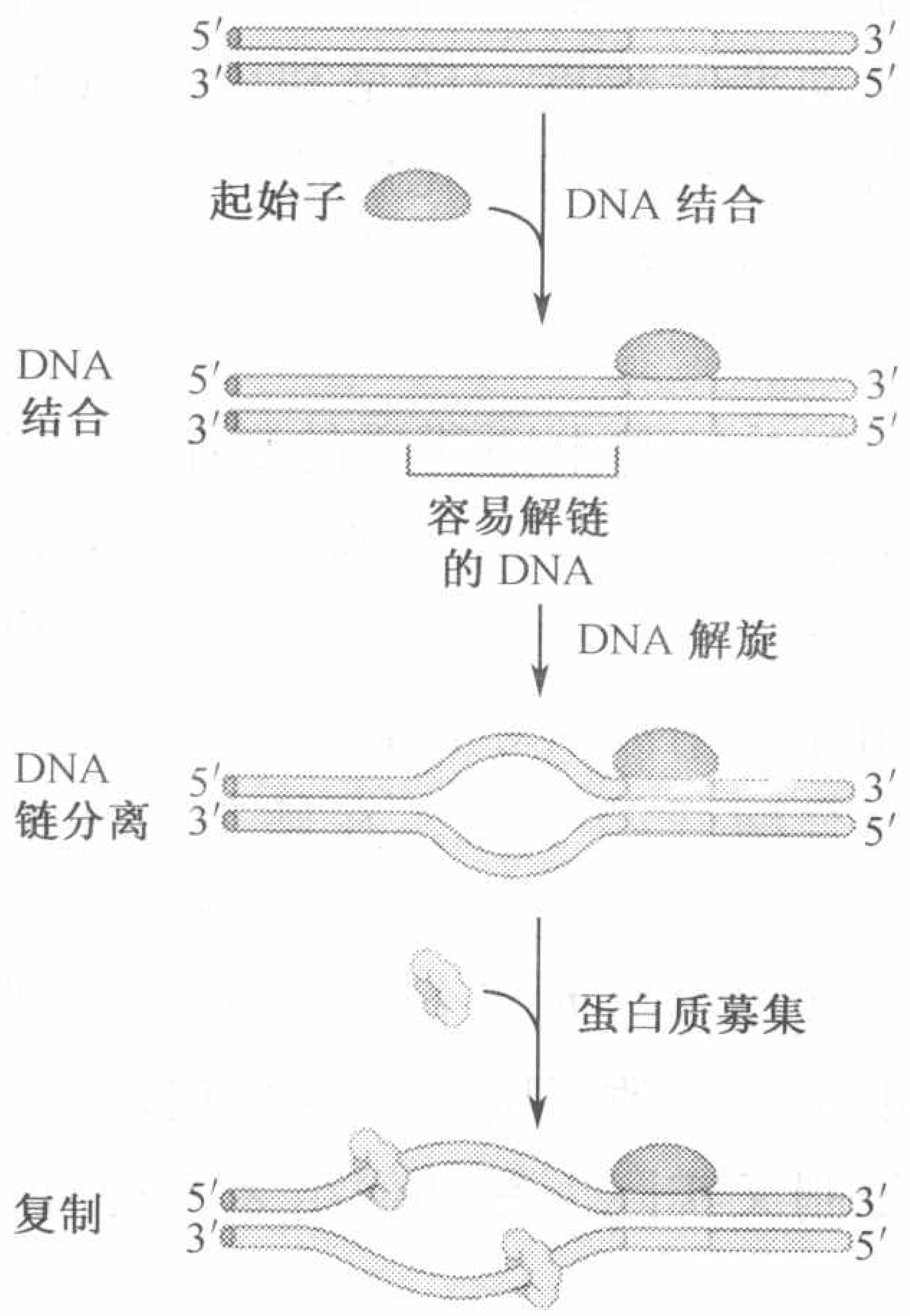


图 8-26　DNA 复制起始期间起始子蛋白的功能。 图示为起始子蛋白 3 种常见的功能：DNA 结合、DNA 链分离以及复制蛋白的募集。(此处图示募集到的蛋白质为 DNA 解旋酶；然而每种起始子蛋白募集到的蛋白质都不相同。)

以 *E. coli* 起始子蛋白 DnaA 为例。DnaA 与 *oriC* 中重复的 9 核苷酸单位元件结合（图 8-25）并受 ATP 的调控。当与 ATP（非 ADP）结合时，DnaA 还与 *oriC* 中重复的 13 核苷酸单位元件所在区域内的 DNA 相互作用。这些额外的相互作用使 13 核苷酸单位元件区域内 20bp 的 DNA 链分离。此解旋 DNA 给其他复制蛋白提供了单链 DNA 模板，用于启始复制中的 RNA 和 DNA 合成步骤（见下文）。

在染色体一个位置上形成单链 DNA 不能满足 DNA 解旋酶和其他复制蛋白的组装，DnaA 还会在复制器所形成的单链 DNA 上聚集其他的包括 DNA 解旋酶的复制蛋白（见下文）。*E. coli* 复制的调控与 DnaA 活性的变化有关，对此将在框 8-7 中讨论。

真核细胞中，起始子是一个被称为起始位点识别复合体（origin recognition complex，ORC）的六蛋白复合体。ORC 的功能在酵母细胞中研究得最清楚。ORC 可识别酵母复制器上被称为 A 元件的保守序列，以及次保守的 B1 元件（图 8-25)。像 DnaA 那样，ORC 结合并水解 ATP。在起始位点上的序列特异性的 DNA 结合需要结合 ATP，ORC 将真核细胞 DNA 解旋酶装载到复制器 DNA 的过程需要水解 ATP 与 DnaA 不同的是，与酵母复制器结合的 ORC 自身并不指导相邻 DNA 链的分离。但是，

ORC 可直接或间接的将其他复制蛋白全部召集到复制器上（见下文）。

蛋白质-蛋白质和蛋白质-DNA 相互作用指导起始过程

起始子与复制器结合后，复制起始中剩下的步骤主要由蛋白质相互作用和序列非依赖性的蛋白质-DNA 相互作用来驱动。最终结果是组装出两个我们前面描述的复制叉机器。为了阐述形成这些蛋白质机器的事件，我们从了解最透彻的 *E. coli* 开始。

起始子（DnaA）结合到 *oriC* 并使 13 核苷酸单位的 DNA 解旋后，单链 DNA 和 DnaA 共同召集双蛋白复合体：DNA 解旋酶、DnaB 以及解旋酶装载器 DnaC（图 8-27），这两种蛋白质在复合体中均为六拷贝。在解旋酶/解旋酶装载器复合体中，DNA 解旋酶保持非活性状态。一旦在起始位点结合单链 DNA，解旋酶装载器就指导它结合的 DNA 解旋酶在单链 DNA 周围进行组装（注意单链 DNA 穿过解旋酶六体蛋白环中央）。这个过程类似于套在引物-模板接头上的滑动 DNA 夹的组装，并且 DnaC 和滑动夹装载器一样是利用 ATP 的 AAA^+ 蛋白（框 8-4）。此过程结束后，解旋酶装载器即解离下来活化解旋酶。两条分离的单链 DNA 链每一条都在起始位点位置上安装一个解旋酶，这两个解旋酶的取向是沿着其所结合的单链 DNA 按 $5'\rightarrow3'$ 极性相向而行。

解旋酶与上面所说的复制叉其他元件之间的蛋白质相互作用指导复制机器其他部分的组装（图 8-27）。解旋酶将 DNA 引物酶召集到 DNA 的起始位点上，导致在起始位点的各条链上合成一条 RNA 引物。DNA 聚合酶Ⅲ全酶通过与引物-模板接头以及解旋酶的相互作用被引导到起始位点上。当全酶出现后，即在 RNA 引物上组装滑动夹，前导链聚合酶开始工作。在解旋酶的作用下，新的单链 DNA 不断暴露并被 SSB 结合，DNA 引物酶合成第一条后随链引物。这些新的引物-模板接头被夹子装载器锁定，此夹子装载器将另外两个滑动夹安放到后随链上。这些夹子被剩下的空闲核心 DNA 聚合酶Ⅲ识别，导致后随链合成的启动。这时，两个复制叉已经组装完成，复制起始结束（两个复制叉实际上是如何组装的还有待定论，见框 8-6）。

真核染色体每一个细胞周期精确地复制一次

如第 7 章所讨论的，真核细胞分裂所需的事件发生在细胞周期的不同时期。染色体 DNA 复制仅发生在细胞周期的 S 期。在此期间，细胞中的所有 DNA 都必须精确地复制一次。染色体任何部分复制的不完整都可造成子染色体之间不适当的连接。互联染色体的分离会引起染色体的断裂或缺失（图 8-28）。DNA 的重复制产生细胞很难修复的损伤。频繁地试图修复这些损伤，导致相关 DNA 的扩增，从而不适当的增加了相关基因的表达。重要调控基因的拷贝数即使增加一个或两个也会导致基因表达、细胞分裂或对环境信号应答的灾难性缺陷。所以，真核细胞每分裂一次，每条染色体中每个碱基对复制且只能复制一次，是极其重要的。

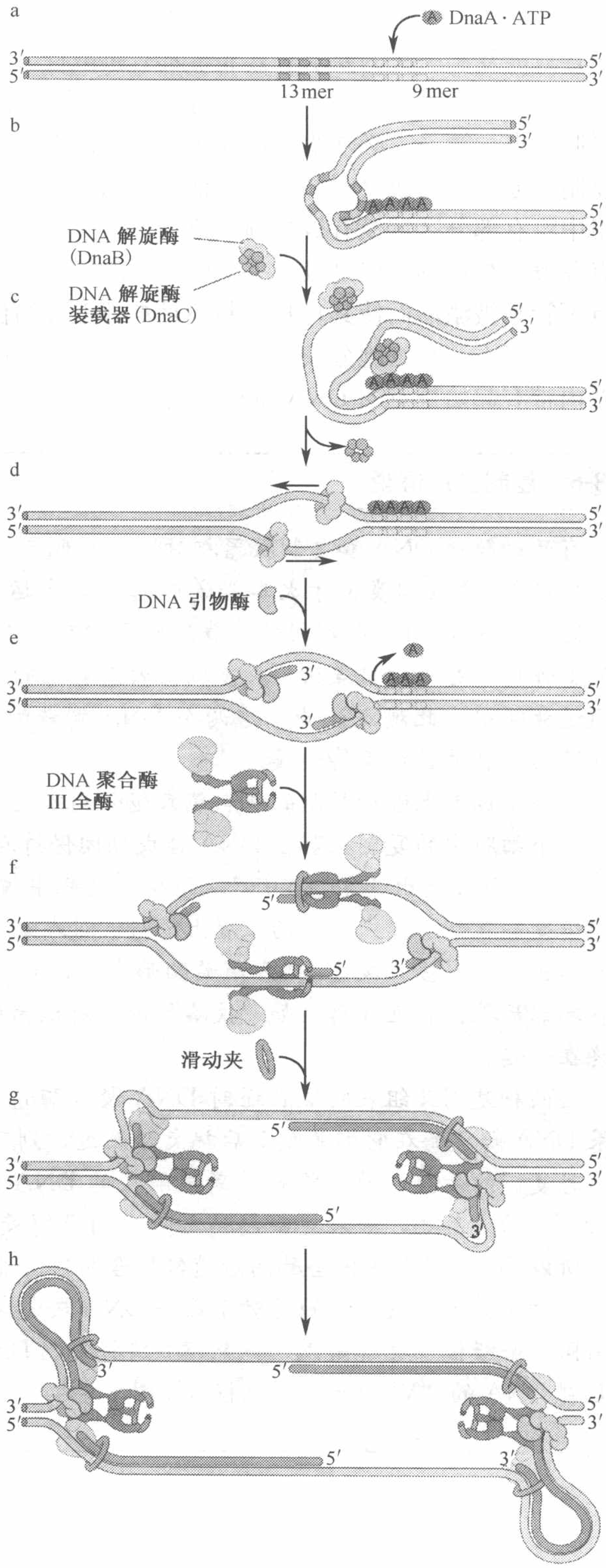


图 8-27 *E. coli* DNA 复制起始的模型。 图示为 *E. coli* 复制起始的主要事件。a. 多个 DnaA-ATP 蛋白结合到 *oriC* 内的 9 核苷酸单位重复序列上。b. DnaA-ATP 与这些序列的结合导致链在 13 核苷酸单位重复处分离。这是由 DnaA-ATP 中的单链 DNA 结合域介导的。c. 与DnaA 结合的 DNA 解旋酶（DnaB）和 DNA 解旋酶装载器（DnaC）与起始位点结合。形成此复合体需要解旋酶装载器中的单链 DNA 结合域以及与 DnaA 的蛋白质相互作用。d. DNA 解旋酶装载器催化 DNA 解旋酶蛋白环的打开和环在单链 DNA 起始位点位置上的套装。DNA 解旋酶的安装导致解旋酶装载器从复制器上脱离下来并激活 DNA 解旋酶。e. 每个 DNA 解旋酶召集一个 DNA 引物酶，DNA 引物酶在各模板上合成一条 RNA 引物。DNA 解旋酶的运动还将所有剩下的与复制器结合的 DnaA 除去。f. 新合成的引物被两个 DNA 聚合酶Ⅲ全酶中的夹子装载器元件识别。滑动夹在每条引物上组装，前导链的合成是由各个全酶中的两个核心 DNA 聚合酶Ⅲ全酶中的一个启动的。g. 当各个 DNA 解旋酶运动约 1000bp 之后，在各自的后随链模板上合成第二条 RNA 引物并装载滑动夹。产生的引物-模板接头被全酶中的第二个 DNA 聚合酶Ⅲ核心酶识别，导致后随链的合成启动。h. 现在各个复制叉上的前导链和后随链合成都已启动，并将持续到模板的末端或遇到邻近复制起始位点产生的另一个复制叉为止。

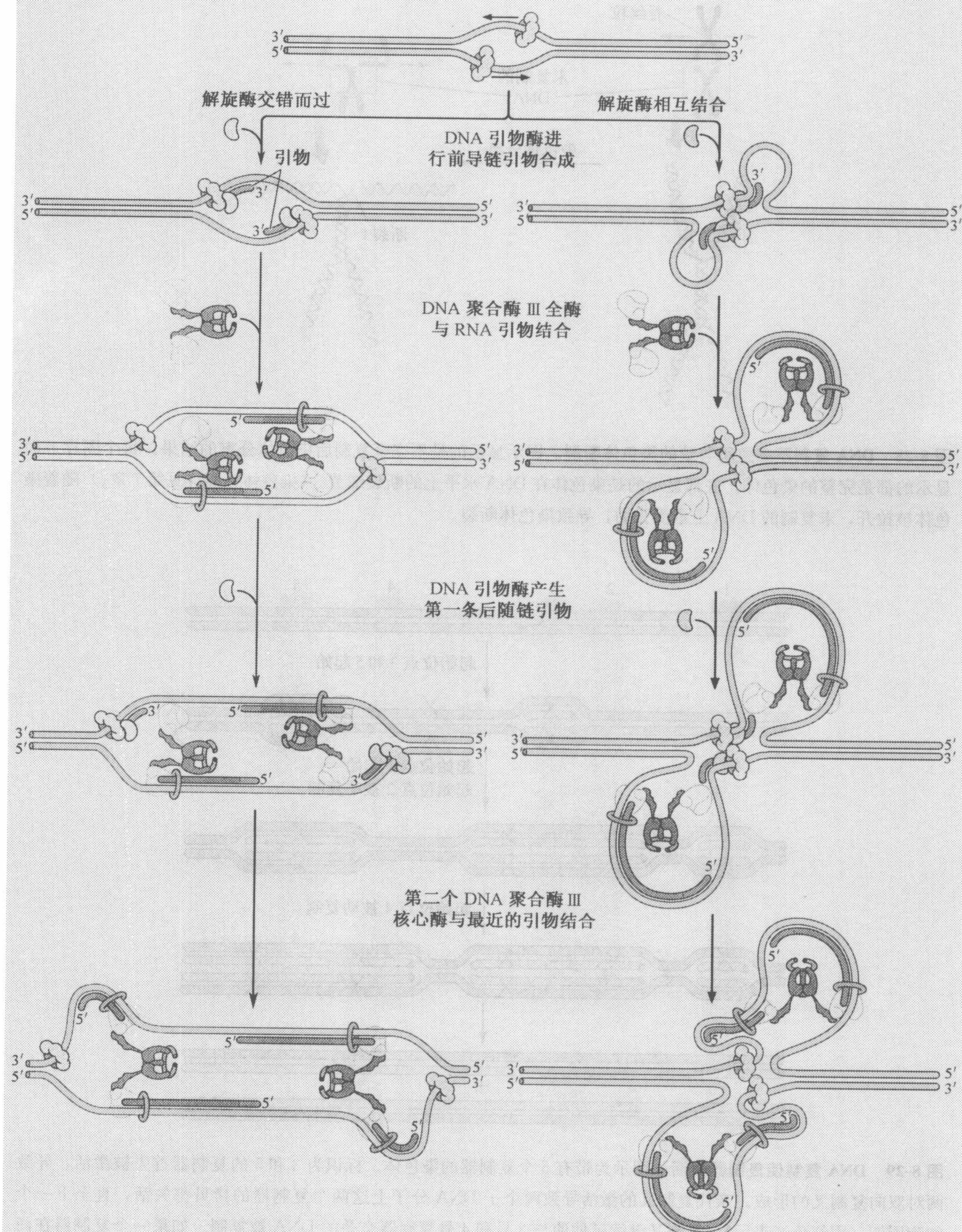
对真核生物染色体来说，由于每条染色体上都有很多个复制起始位点，所以 DNA 复制且只能复制一次的要求对其非常具有挑战性。通常每隔 30kb 就有一个起点，所以即便是一个小的真核染色体可能含有超过 10 个复制起点。人的染色体则含有成千上万个复制起点。第一，必须激活足够多的起始位点，以保证每个 S 期中每条染色体都被完全复制。通常，完成复制并不需要把所有的起始位点都激活，但是如果激活得太少，基因组中将有部分区域没有被复制。第二，虽然在某些细胞分裂轮次中，有些可能的起始位点并没有被用到，但是没有一个已复制的复制起始位点可以重启动。因此，一个起始位点不管是被激活而使其自身复制还是被邻近起始位点产生的复制叉所复制，它都必须失活，直至下一轮细胞分裂开始为止（图8-29）。如果这些条件没有达到，则一个细胞周期内带有起始位点的 DNA 将被复制两次。

框 8-6 复制工厂假说

有两种解释 DNA 和复制机器相对运动的观点（框 8-6 图 1）。一种简单的看法是复制机器沿着 DNA 以类似于火车沿着轨道的方式运动，触及的 DNA 两条链得到复制。此传统观点认为，DNA 解旋酶装载后就立即交错而过，随后各自独立地在两个新的复制叉上作用。另一种观点认为，复制机器是静止的，而 DNA 在运动，这类似于电影胶片通过放映机。此观点认为，这两个 DNA 解旋酶并不是交错而过而是“彼此邂逅”，并在随后复制过程中保持连接。

在静态位置上进行复制的观点越来越受到欢迎。对细菌 DNA 复制的研究清楚地表明，一个细胞中的复制机器在 DNA 合成期间保持在一个位置上。并不是复制机器在运动，而是 DNA 进出“复制工厂”，并在此过程中被复制。类似地，人们观察到在真核细胞中复制在细胞核内分散的位置上进行。对在复制叉上作用的解旋酶的研究也支持静态复制机器。一些六聚体 DNA 解旋酶形成 2 倍六聚体。这说明两个六聚体解旋酶并不是在起始后很快相互分离（如“铁路”模型所认为的那样），而是在整个复制过程中都始终在一起。

这两种复制叉组装的观点还对 DNA 聚合酶Ⅲ全酶复制的 DNA 产生有趣的推理。如果 DNA 解旋酶在它们装载之后就交错而过，则可被 DNA 聚合酶Ⅲ全酶的两个聚合酶同时复制的最接近的链将是最新被解旋的 DNA 的 Waston 链和 Crick 链（框 8-6 图 1，左半边）。反之，如果复制起始之后这两个解旋酶仍连接在一起，因为它们彼此很近，所以 DNA 聚合酶Ⅲ全酶的后随链聚合酶有可能与两条引物化的模板中的任一条结合。这种情况下，最可能的是对于每个 DNA 聚合酶Ⅲ全酶来说，前导链和后随链合成将用同一条模板。也就是说，一个核心酶将复制 DNA 的“Crick”链，而另一个核心酶将复制 DNA 的“Waston”链（框 8-6 图 1，右半边）。



框 8-6 图 1 两种复制机制假说，在左半边，两个 DNA 解旋酶独立工作；在右半边，两个 DNA 解旋酶相互连接。注意在右半边，一个 DNA 聚合酶Ⅲ全酶只用 Waston 链作为模板，而另一个只用 Crick 链作为模板。为简便起见，DNA 聚合酶Ⅲ没有显示连接有 DNA 解旋酶。

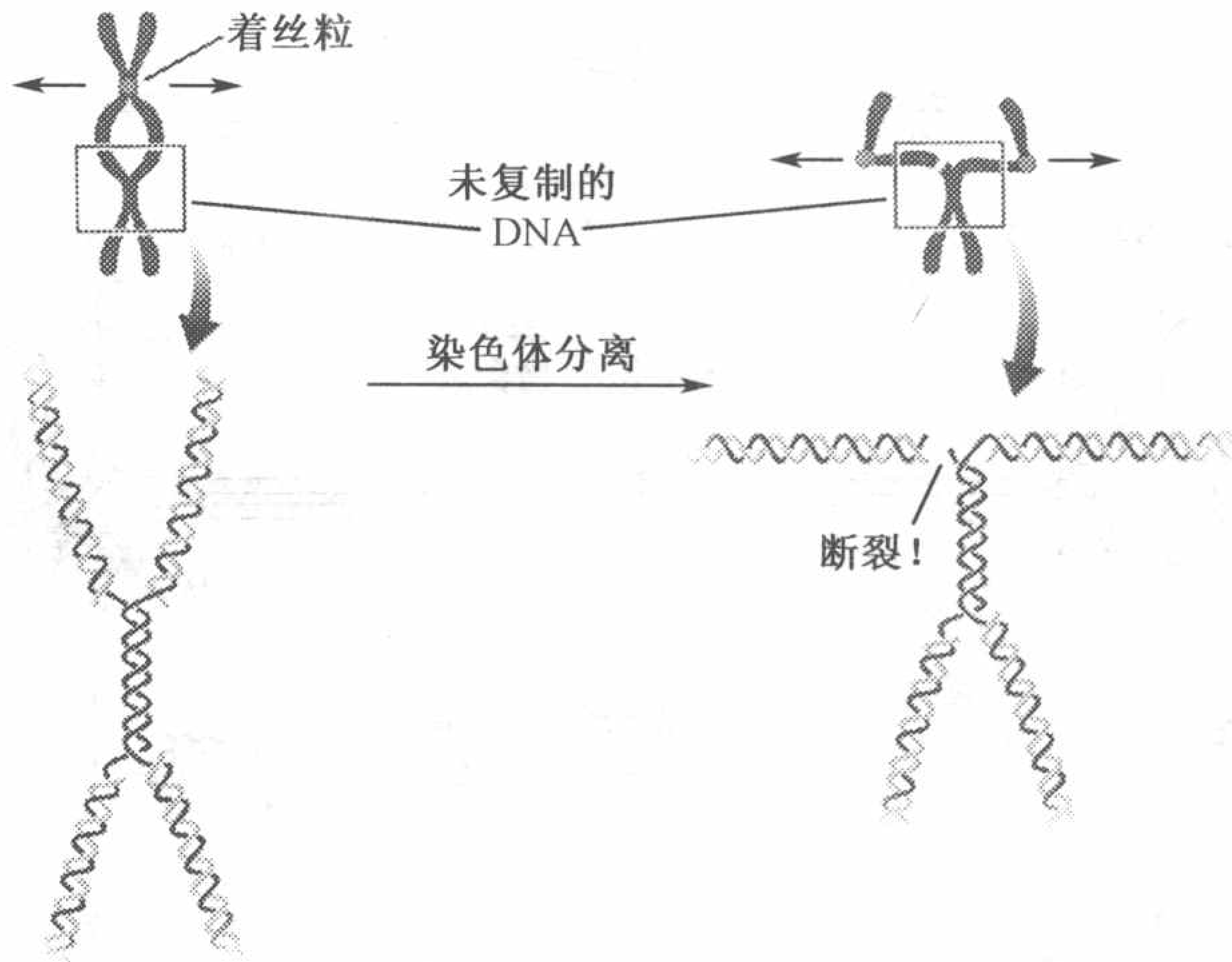


图 8-28　DNA 复制不完整所导致的染色体断裂。图示显示的是不完全复制后染色体分离的结果。每个图片上面显示的都是完整的染色体。底部显示的是染色体在 DNA 水平上的断裂细节。(染色体分离详见第 7 章。)随着染色体被拉开，未复制的 DNA 上承受张力，导致染色体断裂。

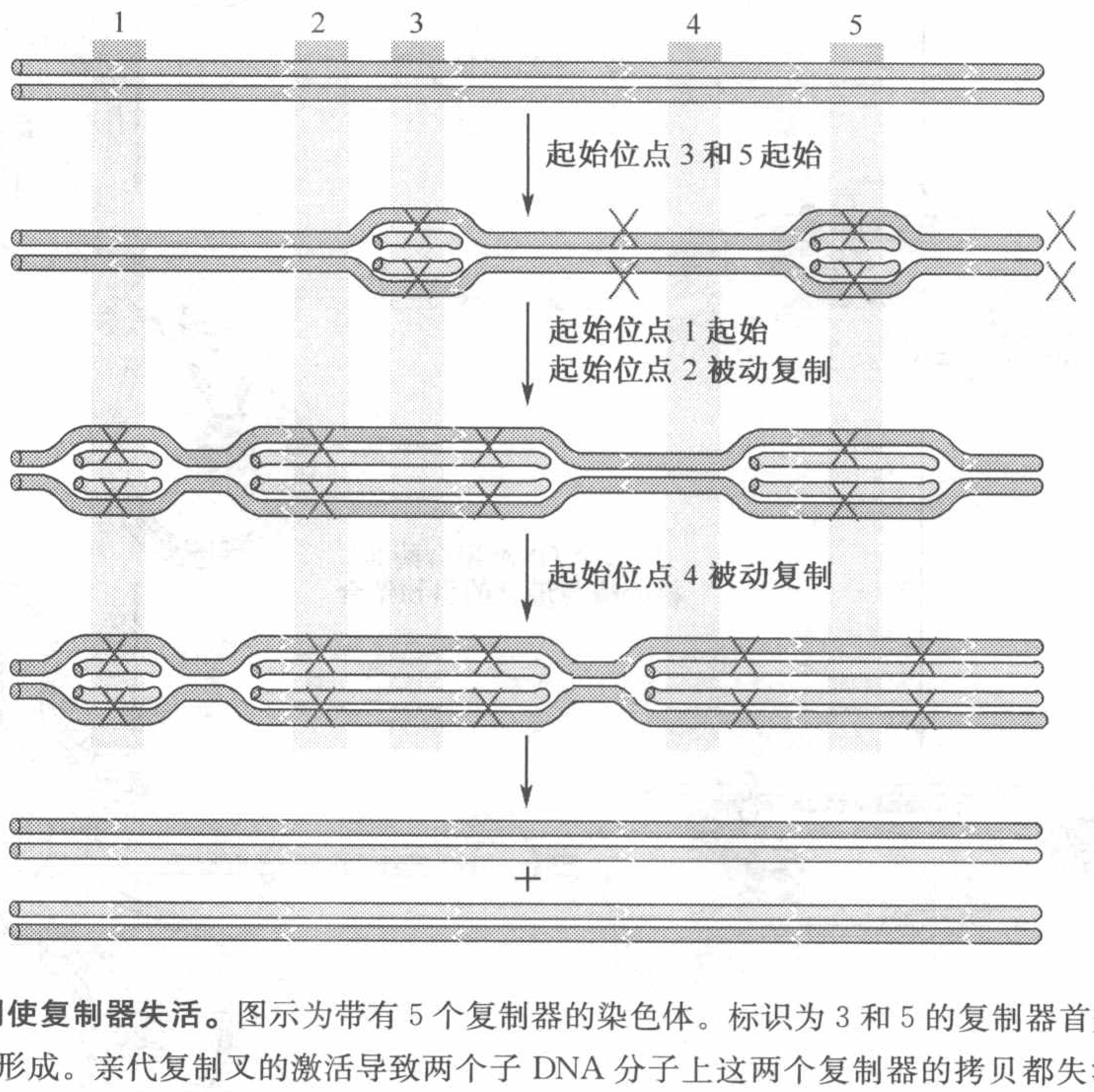


图 8-29　DNA 复制使复制器失活。图示为带有 5 个复制器的染色体。标识为 3 和 5 的复制器首先被激活，导致两对双向复制叉的形成。亲代复制叉的激活导致两个子 DNA 分子上这两个复制器的拷贝都失活，直至下一个细胞周期（用红色×表示）。复制叉继续延伸使与 2 号和 4 号复制器交叠的 DNA 被复制。如果一个复制器在尚未被启动的时候被邻近起始位点过来的复制叉复制，则称其为被动复制。虽然这些复制器并未被启动，但是它们仍然因其 DNA 的复制作用而失活。相反，复制器 1 在启动之前邻近的复制叉并未抵达，它就能正常启动。复制器出现的数目多于完成 DNA 复制所需的复制器的数目，这是一种保证每条染色体都能完成复制的冗余形式。

前复制复合体的形成指导真核细胞中的复制起始

真核细胞中复制起始需要的两个步骤发生在细胞周期的不同时期内（第 7 章）：复制器选择和起始位点的激活。复制器选择是对指导复制起始的序列进行识别的过程，发生于 G_1 期（早于 S 期）。此过程导致基因组中每个复制器上都组装一个多蛋白复合体。起始位点激活只在细胞进入 S 期后发生，使复制器结合的蛋白复合体启动 DNA 的解旋和 DNA 聚合酶的募集。

复制器选择和起始位点激活的分离与原核细胞中的情况不同。原核细胞中复制器 DNA 的识别本质上与 DNA 的解旋以及聚合酶的募集相耦联。如我们下面所看到的那样，在真核细胞中，这两个事件的短暂分离保证了在每个细胞周期中每条染色体仅仅复制一次（细菌细胞对此问题的解决方法不同，见框 8-7）。

复制器选择是由前复制复合体（pre-RC）的形成介导的（图 8-30）。pre-RC 由 4 个独立的蛋白质在复制器中以有序的方式组装而成。pre-RC 形成的第一个步骤是复制器被真核起始子 ORC 识别。ORC 结合上以后，它募集两个解旋酶装载蛋白（Cdc6 和 Cdt1）。ORC 和装载蛋白共同募集真核复制叉解旋酶（Mcm 2-7 复合体）。有趣的是，ORC 和 Cdc6 蛋白与 DnaC 和滑动夹子装载器的亚基一样都是 AAA^+ 家庭成员。对 pre-RC 组装的研究表及这些蛋白质在双链 DNA 周围装载环状 Mcm2-7 复合物过程中具有 ATP 结合和水解活性。pre-RC 的形成并不导致起始位点 DNA 立刻被解旋或者 DNA 聚合酶的募集，而是只有在细胞从细胞周期的 G_1 到达 S 期后，G_1 期形成的 pre-RC 才被激活，并启动复制起始。

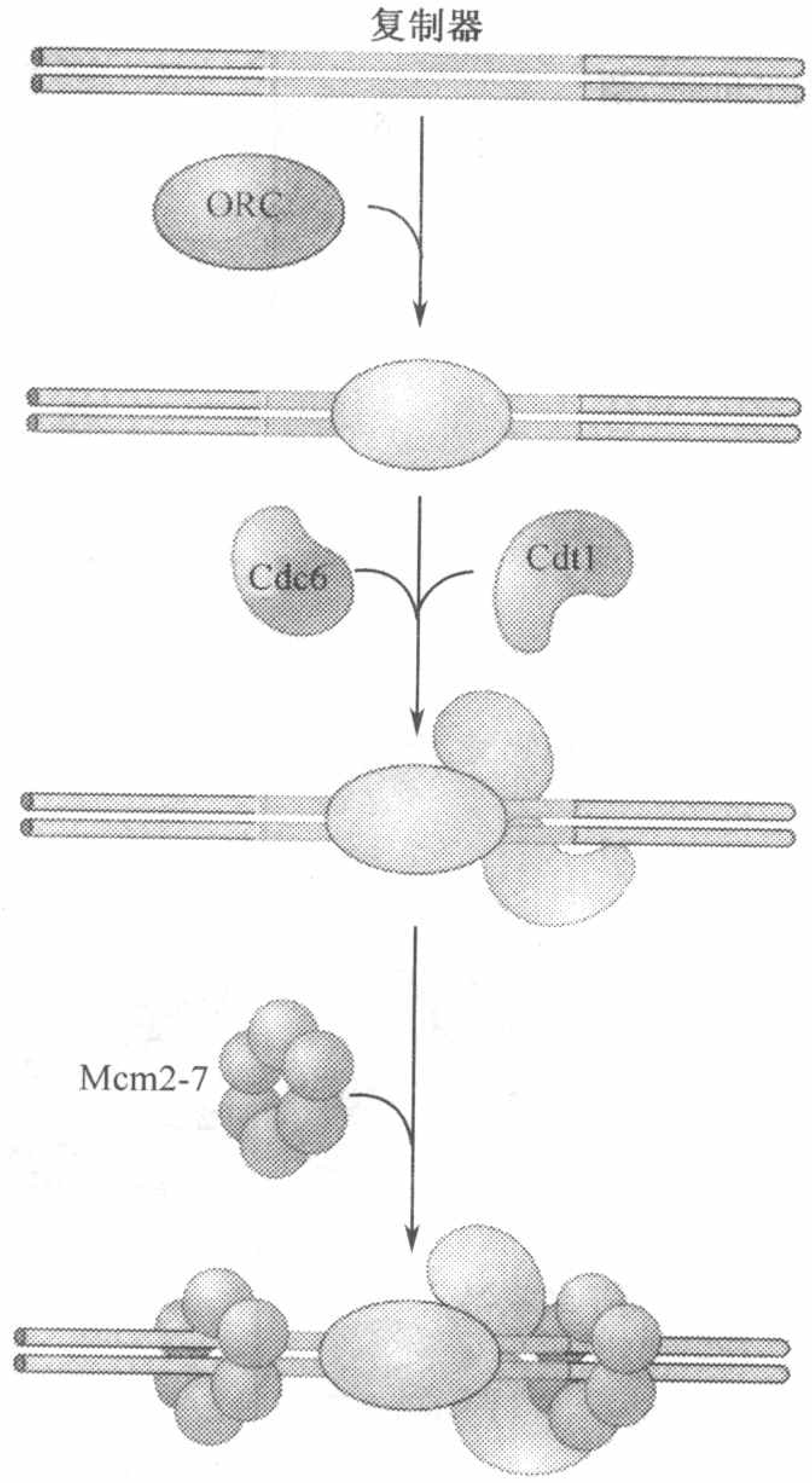


图 8-30　前复制复合体（pre-RC）形成的步骤。 pre-RC 的组装是一个有序的过程，由起始位点识别复合体与复制器的结合引发。ORC 与复制器结合后，至少募集另外两个蛋白：Cdc6 和 Cdt1。这 3 个蛋白质共同作用募集真核 DNA 解旋酶——Mcm2-7 复合体来完成 pre-RC 的形成。

pre-RC 被两种蛋白激酶 Cdk（细胞周期依赖性激酶）和 Ddk（Dbf4 依赖性激酶），（图 8-31）激活，使复制得以启动。激酶是可将磷酸基团共价连接到靶蛋白上的蛋白质（第 5 章）。所有的激酶都在 G_1 期失活，只有在细胞进入 S 期以后才能被激活。当其被激活后，这些激酶以 pre-RC 和其他复制蛋白为靶蛋白。这些蛋白质的磷酸化导致在起始位点上其他复制蛋白的组装以及复制的起始（图 8-31）。这些新的蛋白质包括有 3 种

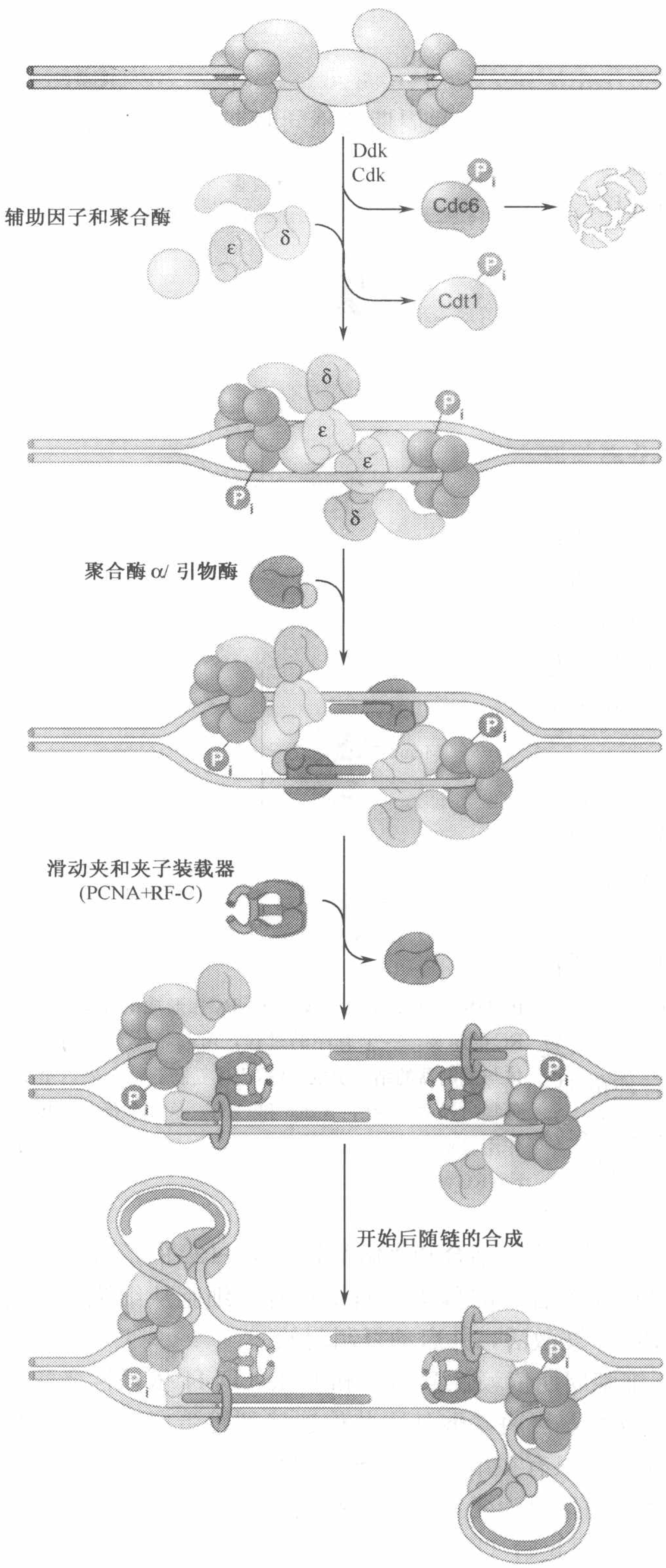


图 8-31　pre-RC 的激活导致真核复制叉的组装。当细胞进入细胞周期的 S 期时，Cdk 和 Ddk 对复制蛋白进行磷酸化以引发复制起始。对导致 DNA 在起始位点上被解旋的事件所知甚少，但是很可能需要 Mcm 复合体的活性从而募集很多辅助因子以及 DNA 聚合酶 δ 和 DNA 聚合酶 ε。DNA 聚合酶 α/引物酶仅在 DNA 聚合酶 δ 和 DNA 聚合酶 ε 之后才被募集。DNA 聚合酶 α/引物酶在起始位点出现之后，合成一段 RNA 引物并进行有限的延伸。产生的引物-模板接头被真核滑动夹装载器（RF-C）识别，并在这些位点上组装滑动夹（PCNA）。DNA 聚合酶 δ 和 DNA 聚合酶 ε 都能识别此引物并开始前导链的合成。DNA 解旋一段时间后，DNA 聚合酶 α/引物酶合成额外的引物，使 DNA 聚合酶 δ 和 DNA 聚合酶 ε 可进行后随链 DNA 合成的起始。此处我们图示的是前导链上的 DNA 聚合酶 δ 和后随链上的 DNA 聚合酶 ε。

DNA 聚合酶及其募集所需的许多其他蛋白质。有趣的是，聚合酶在起始位点的组装是以一定的顺序进行的。DNA 聚合酶 δ 和 DNA 聚合酶 ε 先结合，然后是 DNA 聚合酶 α/引物酶。此顺序保证了 3 种 DNA 聚合酶在起始位点的出现都先于首个 RNA 引物合成（由 DNA 聚合酶 α/引物酶进行）。

在起始位点组装的蛋白质只有部分仍作为真核复制体的一部分继续起作用。除了这 3 种 DNA 聚合酶之外，Mcm 复合体和 DNA 聚合酶募集所需的许多因子也成为复制叉机器的部件。与 *E. coli* DNA 解旋酶装载器（DnaC）类似，其他因子（如 Cdc6 和 Cdt1）在其作用结束之后被释放或被破坏（图 8-31）。

对 pre-RC 形成和激活的调控使每个细胞周期中仅有一轮复制发生

真核细胞如何对数以百计甚至数以千计的复制起始位点进行活性控制，以使得在一个细胞周期中甚至没有一个位点被激活超过一次？答案在于细胞周期蛋白依赖性激酶（Cdk）对 pre-RC 的形成和活化所进行的严谨调控。

Cdk 对于 pre-RC 功能的调控有两个似乎矛盾的作用（图 8-32）：第一，如我们上面所述，它们是激活 pre-RC 以启动 DNA 复制所必需的；第二，Cdk 的活性会抑制新 pre-RC 的形成。

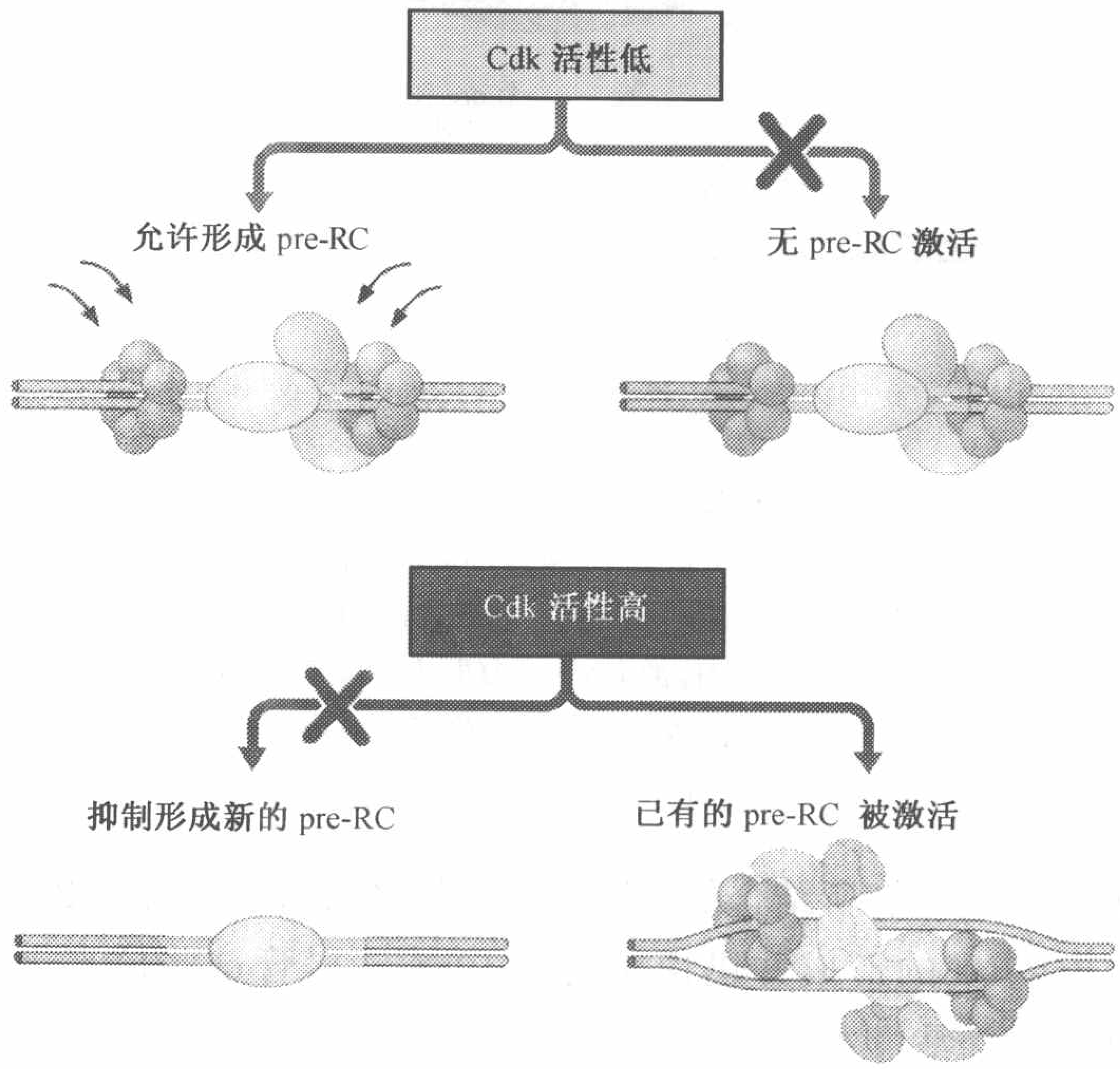


图 8-32 Cdk 对 pre-RC 形成和活化的作用。高 Cdk 活性是已有的 pre-RC 复合物启动 DNA 复制所必需的。Cdk 的这种高活性同时完全抑制了新 pre-RC 复合体的形成；相反，Cdk 的低活性有利于新 pre-RC 的形成但是却不足以通过新形成的 pre-RC 复合体引发 DNA 复制起始。

pre-RC 的功能、Cdk 水平以及细胞周期之间密切的联系保证了真核生物基因组在一个细胞周期内只复制一次（图 8-33）。在 G_1 期内 Cdk 水平低，但是细胞周期的其他

期内（S、G_2 和 M 期）一直存在高水平的 Cdk。因此，在每个细胞周期内，pre-RC 仅仅有一次形成的机会（G_1 期内），而且这些 pre-RC 被激活的机会也仅有一次（实际上在 S、G_2 和 M 期所有的 pre-RC 都被 S 期的复制叉激活或破坏）。

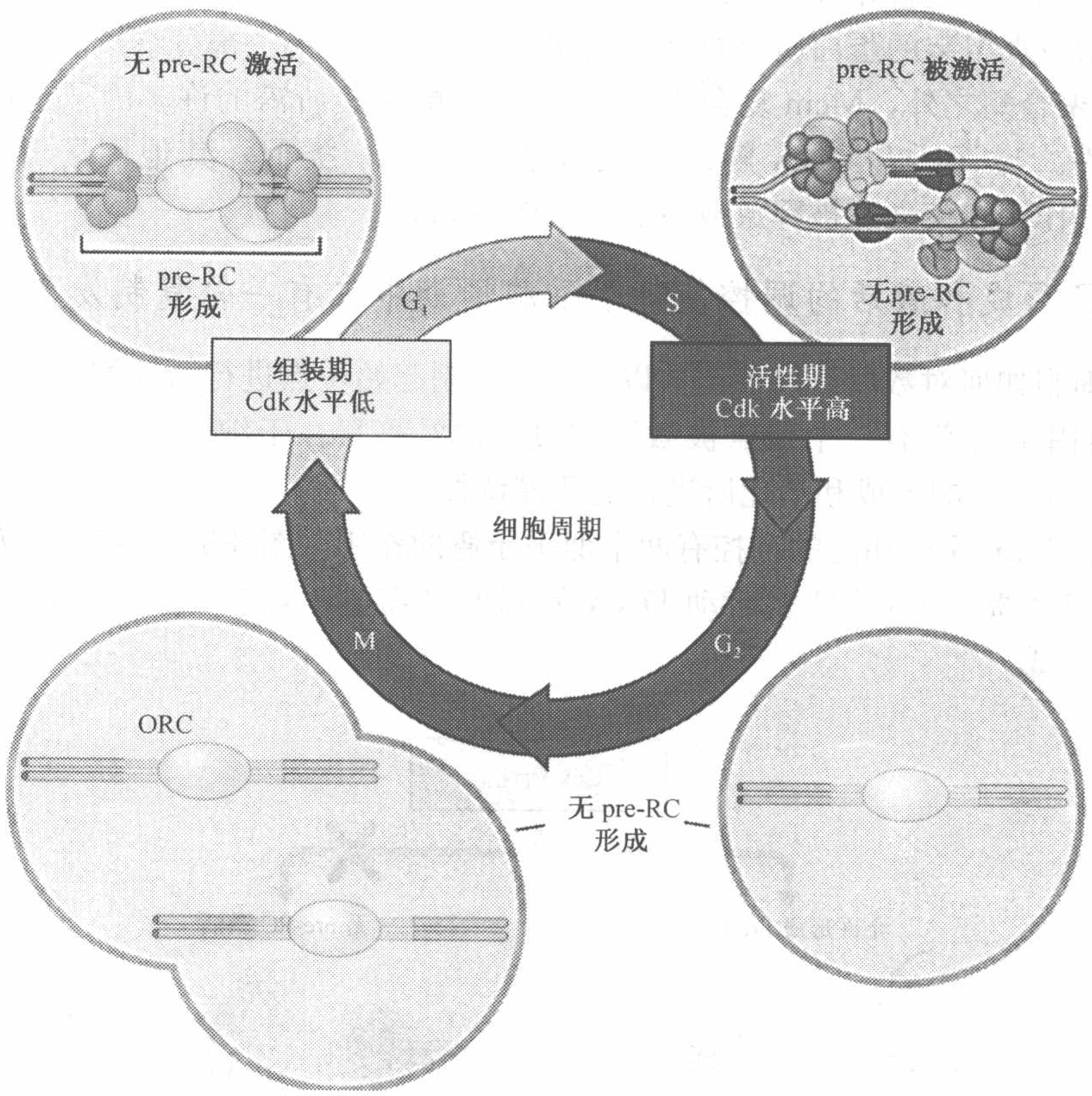


图 8-33 Cdk 活性和 pre-RC 形成的细胞周期调控。在 G_1 期，Cdk 水平低，新的 pre-RC 复合体虽能形成但是却不能被激活。在 S 期，活性水平提高的 Cdk 引发 DNA 复制的起始并阻止任何在刚复制的 DNA 上的新 pre-RC 复合体形成。pre-RC 被用于复制起始之后，就必须解离（注意 pre-RC 至少有一个重要的元件，Mcm 复合体，成为复制叉的部件）。类似地，对与 pre-RC 相连的 DNA 的复制也引起复合体的解体（未显示）。因为 Cdk 在有丝分裂结束之前一直保持高水平，所以在染色体完成分离之前不再形成任何新的 pre-RC 复合体。无新的 pre-RC 复合体形成，复制重新起始是不可能发生的。

pre-RC 在被激活或它们所结合的 DNA 被复制之后解体（图 8-29）。这些暴露的复制器随后用于新 pre-RC 的形成并迅速结合上 ORC。尽管这些位点上存在起始子，但 S、G_2 和 M 期细胞中高活性水平的 Cdk 阻止了 pre-RC 复合体其他成员与 ORC 的结合。只有当细胞将其染色体分离且完成细胞分裂的时候，Cdk 的活性才消失，新的 pre-RC 复合体才能够形成。

真核和原核生物 DNA 复制起始的相似性

现在我们已经描述了在真核细胞和原核细胞中的复制起始，很明显在两种情况下复

制起始有共同的特点。第一步是起始子蛋白对复制器的识别。起始子蛋白和一个或多个解旋酶装载蛋白共同将DNA解旋酶募集到复制器上。解旋酶（和真核细胞中起始位点上其他潜在的蛋白质）产生单链DNA区域，作为RNA引物合成的模板。当引物合成后，复制体的其他元件通过与形成的引物-模板接头相互作用进行组装。

尽管真核细胞和原核细胞中的复制起始相似，但是二者中复制的调节机制是完全不同的。例如，与真核细胞不同，快速分化的细菌细胞在一次细胞循环内多次起始复制。复制调节中最严谨的步骤在真核细胞和原核细胞中也是不同的。真核细胞的对复制的调节集中在MCM解旋酶和DNA最初结合上，而细菌细胞集中在DnaA起始蛋白与DNA的结合上（框8-7）。

框8-7 *E. coli* 复制受DnaA-ATP水平和SeqA的调控

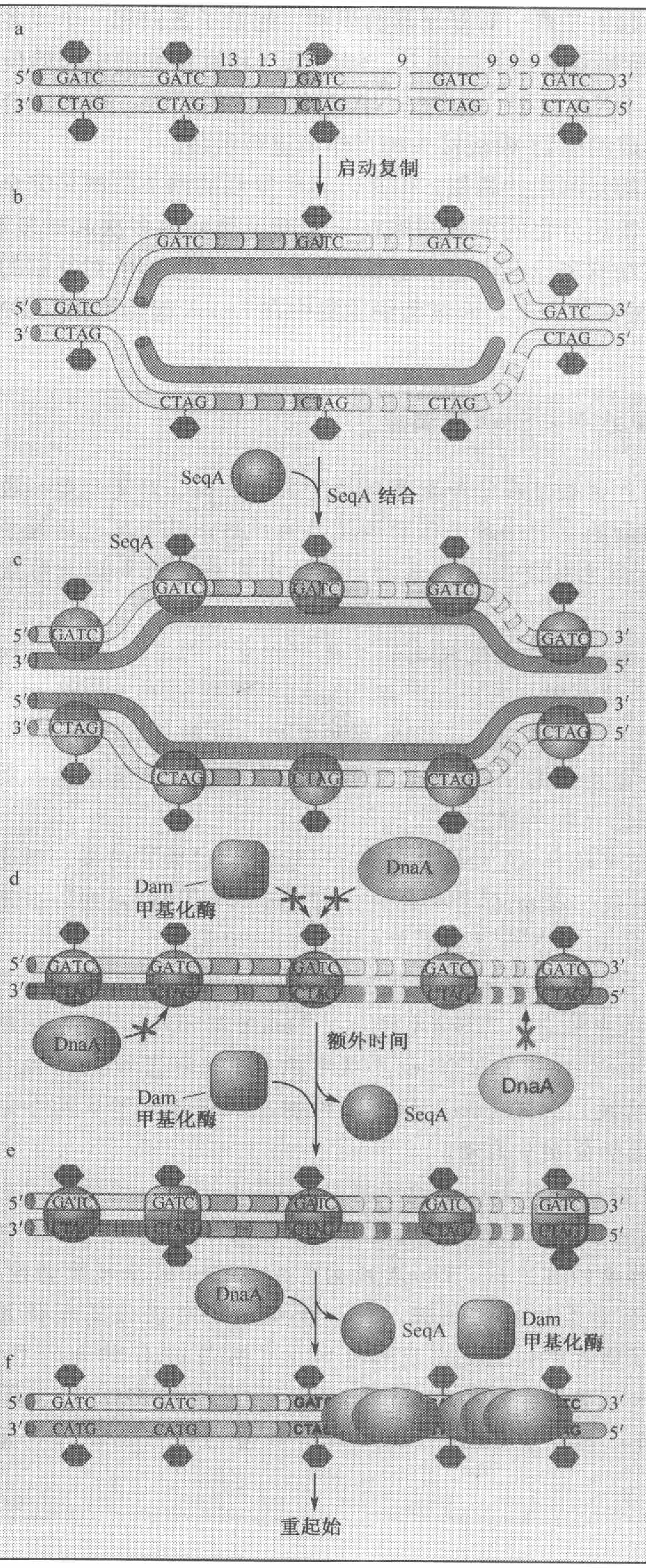
在所有生物中，为了保证染色体数目和细胞数量保持适当的平衡，对复制起始进行严格控制十分重要。虽然在真核细胞中对这种平衡的调控最为严格，*E. coli* 也通过禁止新近启动的起点重新启动而防止染色体复制出现意外。有几个不同的机制用来防止在 *oriC* 上重新启动后的快速复制。

有一个方法可检测DNA复制前后甲基化状况的变化（框8-7图1）。*E. coli* 细胞中，一种称作Dam甲基转移酶的酶将甲基基团加到每个GATC序列的A上（注意此序列是回文结构）。通常基因组在GATC序列上是完全甲基化的。这种情况在GATC序列复制之后就改变了。因为在新合成的DNA链中，A残基尚未甲基化，所以这些刚刚被复制的位点只有一条链被甲基化（即半甲基化）。

新复制 *oriC* 的半甲基化状态可被SeqA检测到。SeqA与GATC紧密结合，但这仅发生在GTAC序列半甲基化的时候。在 *oriC* 紧邻的地方有大量的GTAC序列。当复制起始后，SeqA即在这些位点被Dam甲基化酶完全甲基化之前与之结合。

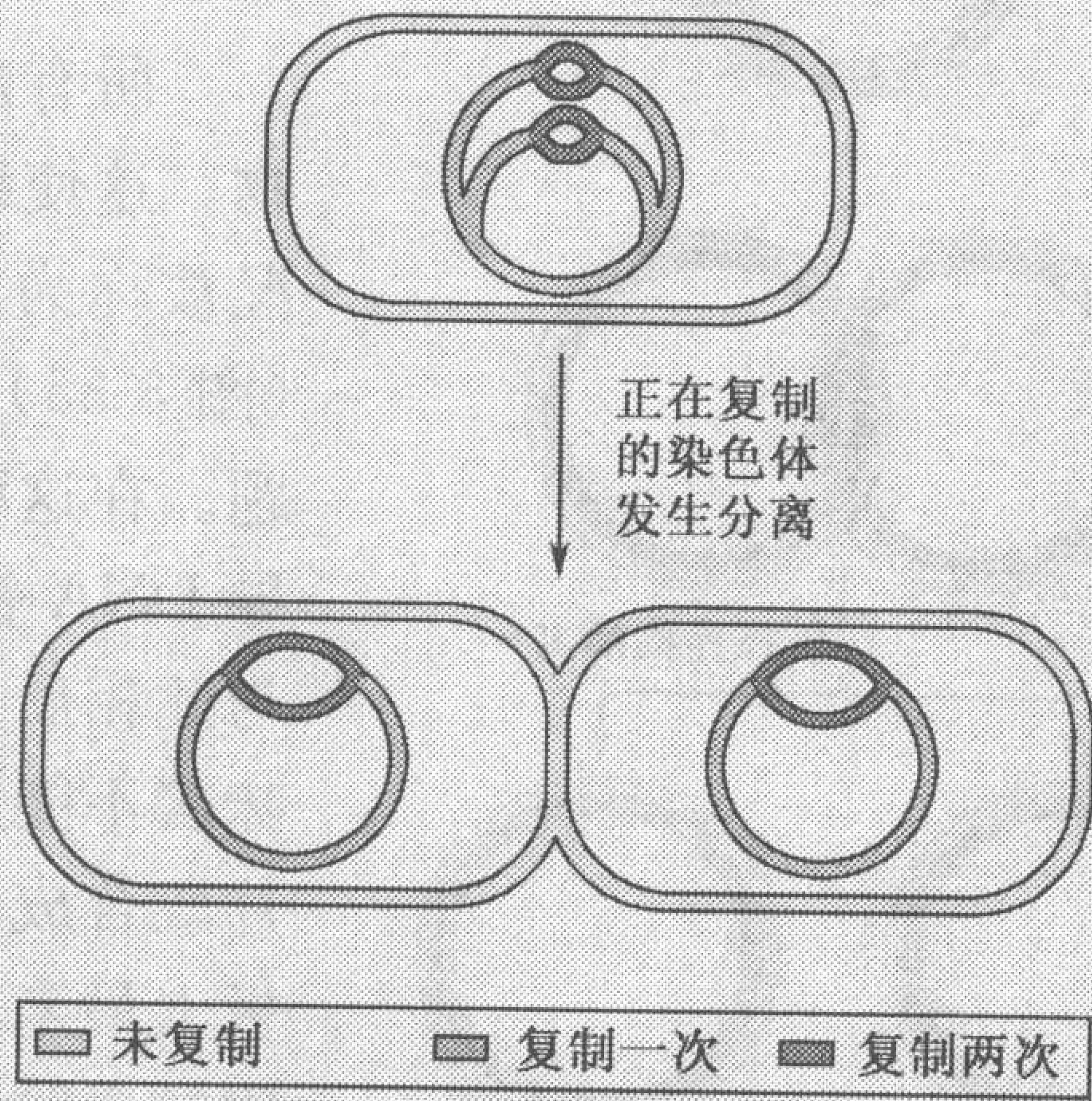
SeqA的结合有两个结果。首先，它极大地降低了GATC结合部位处的甲基化速率；其次，当与这些 *oriC* 邻近位点结合时，SeqA阻止了DnaA与 *oriC* 的结合和结合所导致的新一轮的复制。因此，*oriC* 附近GATC位点从甲基化到半甲基化的转化（此事件是复制从 *oriC* 起始的直接结果）导致DnaA结合受抑制，从而阻止了从两个新合成出的子拷贝的 *oriC* 上进行快速的复制重启动。

其他针对DnaA的机制也可抑制在新合成出的子拷贝 *oriC* 上进行快速的重启。如上所述，只有与ATP结合的DnaA才能指导复制的起始；但在起始过程中，ATP转化成ADP。因此，经过一轮复制起始的过程后，DnaA成为失活状态而无法被重新使用。将结合的ADP换成ATP是一个非常缓慢的过程，进一步阻滞了可促进复制作用的DnaA-ATP结合物的积累。*oriC* 附近序列的复制进程也减少了可与 *oriC* 结合的DnaA的数量。在 *oriC* 外有300多个9核苷酸单位的结合位点（DnaA还在许多启动子上作为转录调节子），当它们被复制时，此数量加倍。DnaA结合位点的增加使得可利用的DnaA数量降低。



框 8-7 图 1 SeqA 与半甲基化的 DNA 结合抑制了新近复制的子代复制起始位点的重起始。 a. 在 DNA 复制之前，*E. coli* 基因组上所有的 GATC 序列都是两条链被甲基化的（“完全”甲基化）。图中甲基基团用红色的六边形表示。b. DNA 复制使这些位点变成半甲基化状态（只有一条 DNA 链被甲基化）。c. 半甲基化的 GATC 序列很快被 SeqA 结合。d. 结合上的SeqA蛋白抑制这些序列被完全甲基化以及 DnaA 蛋白对 *oriC* 的结合。（为了简便起见，d、e、f 部分只图示两条子分子中的一条。）e. 当 SeqA 偶尔从 GATC 位点上脱离时，序列即被 DNA 甲基转移酶完全甲基化，防止了 SeqA 的重新结合。f. 当 GATC 被完全甲基化之后，DnaA 能进行结合并指导新一轮子代 *oriC* 复制器的复制。

这些机制共同使 *E. coli* 在 *oriC* 的新拷贝上启动复制的能力快速显著地降低。虽然这些机制防止了快速的重起始，但是此种抑制却并不需要一直持续到细胞分裂完成之时。事实上，为了使 *E. coli* 细胞用最大速度进行分裂，*oriC* 的子拷贝必须在上一轮复制完成之前就启动复制。这是因为 *E. coli* 细胞每 20min 分裂一次，但是 *E. coli* 基因组复制一次就需要 40 多分钟。因此，在快速生长条件下，*E. coli* 细胞在上一轮复制完成之前就要重新启动一次，有时候是两次（框 8-7 图 2）。即使在这样快速的生长条件下，每轮细胞分裂中复制起始也不超过一次，因此，每一轮细胞分裂就只有一轮自 *oriC* 的复制起始。



框 8-7 图 2　在快速生长的细胞中复制起始位点的重起始先于细胞分裂。为了使基因组在每一轮细胞分裂前完全复制，细菌细胞必须在细胞分裂完成之前在其单起始位点上经常进行 DNA 复制的起始。这意味着分离到子细胞中的染色体仍在进行活跃的复制。这与真核细胞不同，真核细胞中在所有的 DNA 复制完成之前不进行染色体的分离。

结束复制

DNA 复制的完成需要一系列特殊的事件。对于环形和线性染色体来说，这些事件是不同的。对于环形染色体，常规的复制叉机器能复制整个分子，但是产生的子代分子之间是相互拓扑连接的。相反，线性染色体最末端处的复制不能由我们目前所讨论的复制叉机器来完成。因此，含有线性染色体的生物发展出一套新的策略以克服这种末端复制的问题。

子代 DNA 分子的分离需要拓扑异构酶 Ⅱ

环形染色体复制完成之后，产生的子代 DNA 分子如铰链般连接在一起（图 8-34）。铰链是描述两个相连的圆环（类似于链中的连环）的一般性术语。为了使这些染色体分离到子细胞中去，两个环形 DNA 分子必须相互脱离。此分离是由拓扑异构酶 Ⅱ 作用完成的。如我们在第 6 章所讨论的，这些酶具有断裂一条双链 DNA 分子并使第二条双链 DNA 分子通过此缺口的能力。这样，拓扑异构酶 Ⅱ 催化了两个子代 DNA 分子中的一个断裂并使另一个子代分子通过这个缺口。此反应使两个子代染色体解开铰环，并使它们分配到两个分离的细胞中。

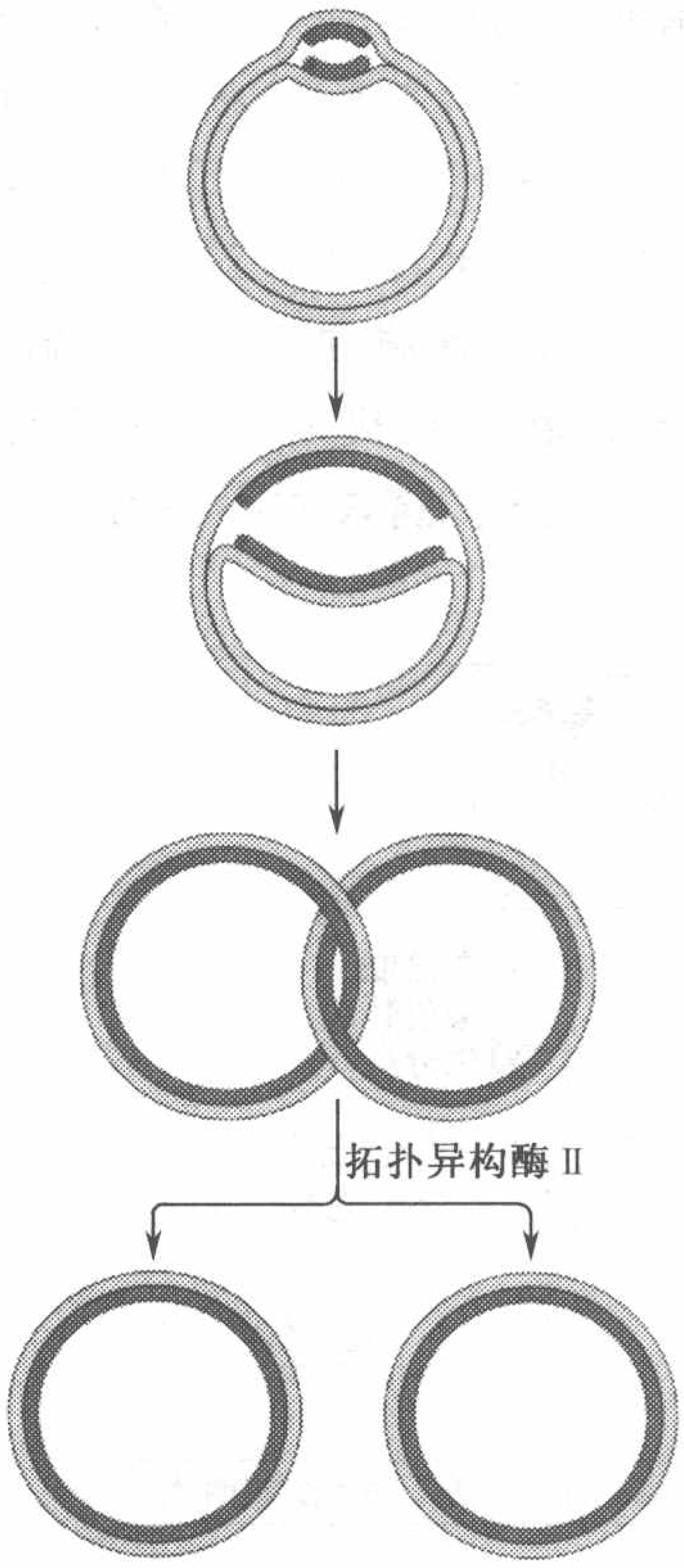


图 8-34 拓扑异构酶Ⅱ催化复制产物的解环。当一个环形 DNA 复制后，产生的子代 DNA 分子仍保持相互连接。DNA 拓扑异构酶Ⅱ能够有效地使 DNA 环分开（或解环）。

虽然此活性对环形染色体的分离的重要性是毋庸置疑的，但是拓扑异构酶Ⅱ的活性对于大的线形分子分离也很重要。尽管线形分子复制之后本身并无拓扑学连接，但是体积巨大的真核生物染色体必须将 DNA 折叠成环并与蛋白质支架连接（第 7 章）。这种连接导致在两个子代染色体分离时会遇到很多与环形染色体分离同样的问题。

后随链的合成不能复制线性染色体的最末端

所有新的 DNA 合成启动都需要一条 RNA 引物，这使线性染色体末端的复制成为难题。这种情况被称为末端复制问题（end replication problem)（图 8-35）。对前导链模板复制来说不存在这种问题。在这种情况下，一条内部的 RNA 引物即可指导 DNA 链的起始并能延伸到其模板的 5′最末端。与此相反，后随链合成的完成需要多条引物，这意味着不能产生一条完整拷贝的模板。即使是在冈崎片段合成的最后一个 RNA 引物的末端与后随链模板上最末端的碱基对退火的情况下，当此 RNA 引物被去除以后，在染色体末端仍将有一小段未复制的单链 DNA，这意味着每一轮 DNA 复制都将出现两条子代 DNA 分子中的一条被截短的情况。很显然，这种情况将破坏遗传物质从一代向另一代的完整传递。最终，染色体末端的基因将会丢失。

细胞解决末端复制问题有各种各样的方法。一种解决方法是用蛋白质代替 RNA 作为每个染色体末端最后一个冈崎片段的引物（图 8-36）。在这种情况下，“引物蛋白”与后随链模板结合并用一个氨基酸来提供—OH，以代替正常情况下 RNA 引物提供的 3′-OH。通过与最后的后随链结合，引物蛋白与染色体的 5′端形成共价连接。这种在末端连接复制蛋白质的情况，在某些具线性染色体的细菌（多数细菌是环形染色体）的染色体末端以及在某些具线性染色体的细菌病毒和动物病毒的染色体末端都有发现。

多数真核细胞使用完全不同的方法来复制其染色体末端。如我们在第 7 章所述，真核生物染色体的末端称为端粒，它们通常由首尾相接的富含 TG 的重复 DNA 序列构成。例如，人端粒含有很多首尾相接的重复序列 5′-TTAGGG-3′。虽然这种重复序列很多都是双链的，但是每条染色体的 3′端都位于其 5′端之外成为单链 DNA。这种独特的结构可作为新的复制起始位点以弥补末端复制问题。此起始位点所相互作用的蛋白质与真核生物其他起始位点的作用蛋白不同，它募集一个称作端粒酶的特殊蛋白来行使功能。

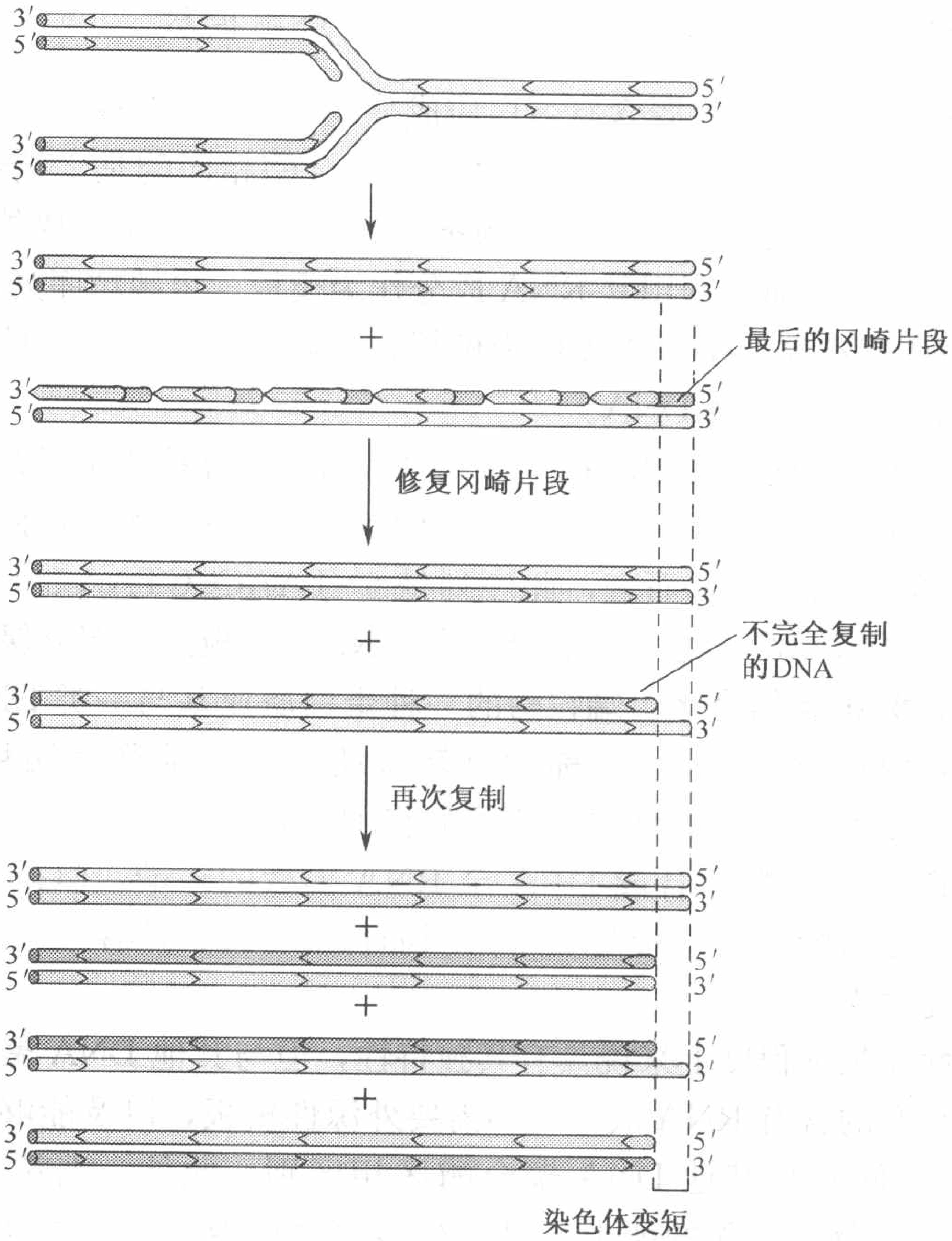


图 8-35 末端复制问题。当后随链复制机器到达染色体末端时，有时引物酶没有足够的空间去合成新的 RNA 引物。这导致复制的不完整和后随链 DNA 产物上 3′端形成一小段单链 DNA。当此 DNA 产物在下一轮被复制的时候，两个产物中的一个将变短，在上一轮复制中没有被复制的区域将丢失。

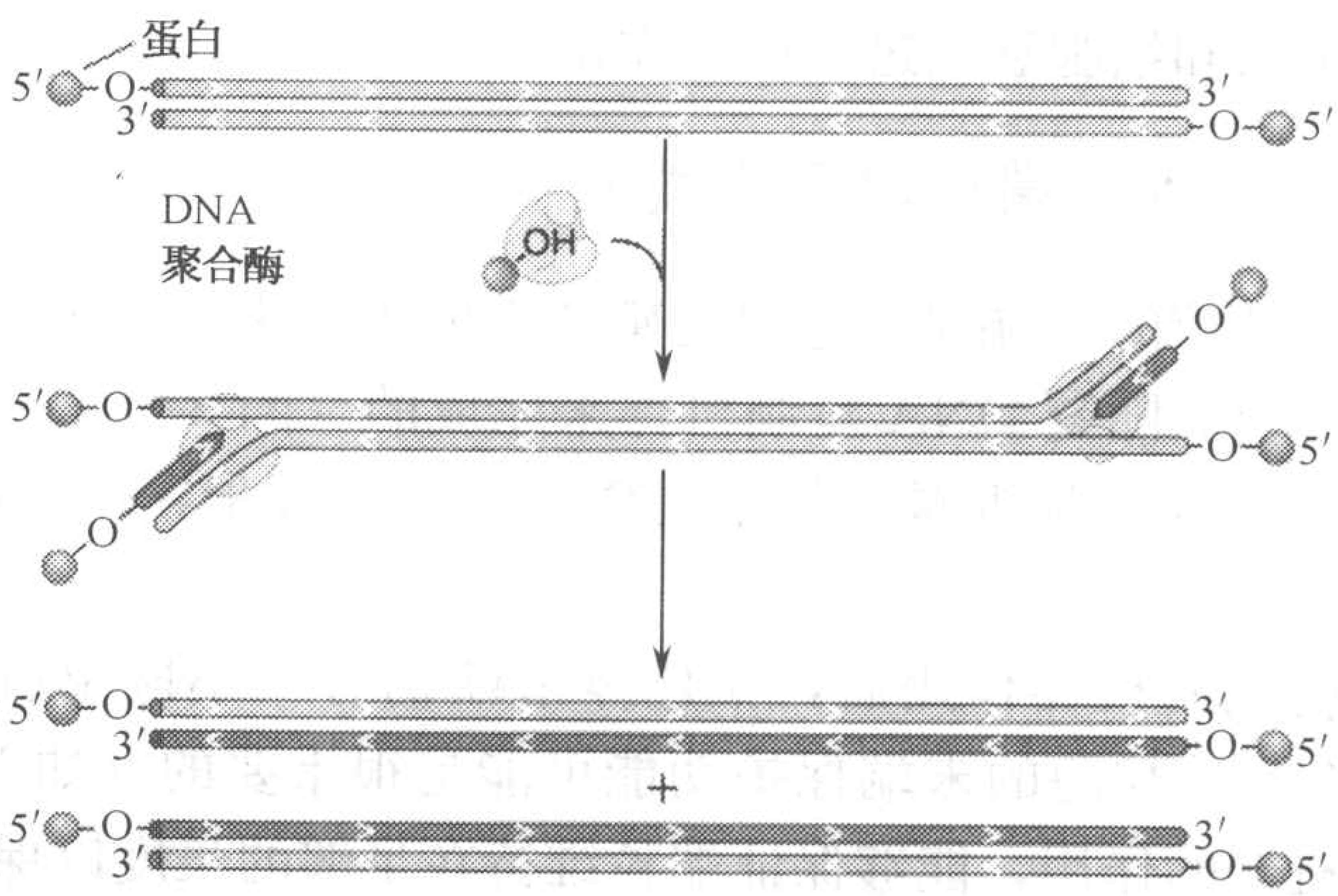


图 8-36 蛋白质引物是解决末端复制问题的一个办法。通过与 DNA 聚合酶和模板 3′端的结合，蛋白质提供具引物作用的羟基基团以起始 DNA 的合成。在所示的例子中，蛋白质如在很多病毒中所见到的那样成为 DNA 合成的引物。对于较长的 DNA 分子来说，此方法结合传统的起始位点处的复制方法来使染色体进行复制。

端粒酶是一种新型的 DNA 聚合酶，它不需要外源模板

端粒酶是一种特殊的酶，它既含有蛋白质成分也含有 RNA 成分（所以这是一个核糖核蛋白的例子，见第 5 章）。与其他所有 DNA 聚合酶相同的是，端粒酶用于延伸其 DNA 底物的 3′端。但与大多数 DNA 聚合酶不同的是，端粒酶不需要外源性 DNA 模板来指导新 dNTP 的添加，而是利用其 RNA 成分作为模板，将端粒序列添加到染色体末端的 3′端。端粒酶利用其自身的 RNA 作为模板，特异性地延伸特定单链 DNA 序列的 3′-OH。新合成的 DNA 为单链 DNA。

端粒酶不同寻常的功能的关键所在是该酶的 RNA 组分称为端粒酶 RNA 或 TER，在不同物种中，TZR 大小从 150bp 到 1300bp 不等。在所有生物中，此 RNA 序列含有 1.5 拷贝的完整端粒序列（对人而言，此序列是 5′-TAACCCTAA-3′），能与端粒 3′端的单链 DNA 退火（图 8-37）。退火发生的方式是 RNA 模板的一部分保持单链，形成可被端粒酶作用的引物-模板结合区。端粒酶的一种蛋白质亚基与一类 DNA 聚合酶相关，这种 DNA 聚合酶使用 RNA 模板，称为反转录酶（此亚基称为端粒酶反转录酶或 TERT）如我们将在第 11 章所看到的，这些酶将 RNA 反转录成 DNA，而不是较常规的 DNA 转录成 RNA。端粒酶合成 DNA 至 RNA 模板的末端，但是不能继续复制此 RNA 之外的序列。此时 RNA 模板与 DNA 产物解离，再次与端粒上的最后 3 个核苷酸退火，然后重复此过程。

端粒酶的特性在某种程度上来说具有其独特性，但与其他 DNA 聚合酶仍是非常类似的。端粒酶所具有的含有 RNA 成分、不需要外源性模板，以及能够使用单纯的单链 DNA 底物的特点，使其与其他 DNA 聚合酶区相区别。此外，端粒酶还必须有将其 RNA 模板从 DNA 产物上去除的能力，以进行多轮模板指导下的重复合成。通常，这意味着端粒酶有 RNA-DNA 解旋酶活性。另一方面，与其他所有的 DNA 聚合酶相同，端粒酶也需要模板来指导核苷酸的添加，只能延长 DNA 的 3′端，使用同样的核苷酸前体，具有持续延伸能力，每次与 DNA 底物结合后可添加很多重复序列。框 8-8 讨论了端粒酶在调节细胞生长和细胞衰老过程中的作用。

端粒酶通过延伸染色体 3′端解决了末端复制问题

当端粒酶作用于端粒的 3′端时，它仅能延伸染色体的这一末端。那么，5′端是怎样延伸的呢？此工作是由后随链 DNA 复制机器来完成的（图 8-38）。通过提供一个延伸的 3′端，端粒酶为后随链复制机器的作用提供了额外的模板，使后随链复制机器可延长 DNA 5′端。

即便是在后随链复制机器作用后，染色体末端仍有一小段单链 DNA 区域。事实上，3′凸出端的存在对端粒酶的末端保护功能可能是很重要的（如下所述）。然而，端粒酶和后随链复制机器的作用，能够保证端粒维持在足够的长度以保护染色体末端不会变得太短。由于端粒 DNA 的重复性以及非蛋白编码性，细胞可以允许不同长度的端粒存在。

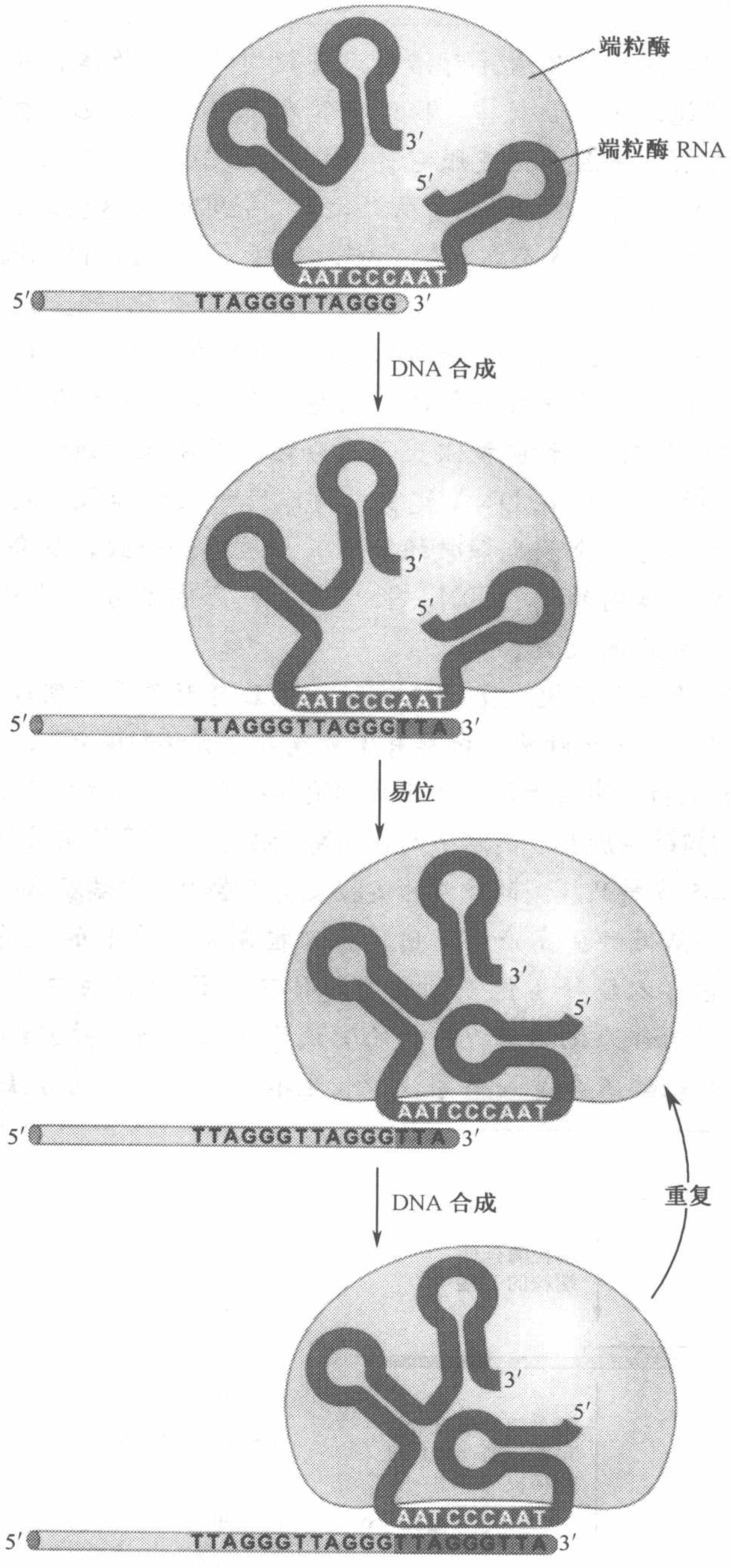


图 8-37 端粒酶进行的端粒复制。端粒酶用其 RNA 成分与端粒单链 DNA 区域的 3′端退火。随后用其反转录活性合成 DNA 至 RNA 模板末端。这时端粒酶将 RNA 从 DNA 产物上移去并重新结合到端粒的末端，重复上面的过程。

框 8-8 衰老、癌症和端粒假说

所有生物终究都要死亡。不管是存活数天或数周的小生物体，还是平均寿命以年计的人类，生物体都不能逃脱死亡规律，因此研究人员（以及其他人员）长期研究细胞复制次数的限度以期理解它并且试图克服它并且能最终“长生不老”。

当研究者发明了体外单个细胞的培养方法之后，他们认为这些细胞是长生不朽的。这就暗示死亡是一个对整个有机体而非细胞而言的难题。Leonard Hayflick 对培养细胞的分化研究推翻了该假说。他发现，在分离状态下，细胞只能分化一定次数。有趣的是，Hayflick 的研究表明细胞分化的次数是细胞寿命的标志，即现在众所周知的 Hayflick 界限。

Hayflick 的研究使人产生这样的想法：细胞含有一个内在的倒数计时钟限定细胞分化的次数。当这个钟到零时，细胞被阻止再次分化。多年来对该倒计时钟签定未果。然而，随着对端粒的特性以及它在 DNA 复制中的作用的认识加深，端粒显而易见可以作为继分化倒计时钟之后另一个长期探索的目标。与该想法一致，从年轻人体内分离的端粒 DNA 比来自年长者体内的端粒 DNA 长。该发现产生了另一种假说，端粒 DNA 的长度限制了细胞可以分化的次数。

尽管此观念仍然是一个假说，支持端粒和细胞衰老有关系的实验证据正在积累。例如，为了使此假说可行，正常细胞应该具有少量或者没有端粒酶活性。否则当这些细胞变短时他们能继续延长端粒。事实上，许多正常细胞具有有限的端粒酶活性。反之，像干细胞和肿瘤细胞能持续增殖的细胞具有高水平的端粒酶活性。对培养基中癌细胞的研究显示他们可以无限分化。第二个支持此模型的重要的实验表明正常细胞中端粒酶的表达使细胞不死。

癌细胞中端粒酶活性的提高产生了用端粒来控制无限生长的细胞生长特性的假说。如果该假说正确，它可以解释为何多细胞有机体不允许所有细胞具有端粒酶活性。确实，寻找端粒酶抑制剂作为化学治疗药物的尝试已付出良多。癌症细胞中端粒酶活性的提高同时暗示激活端粒酶不会是一个寻找“长生不老”的明智的方法。

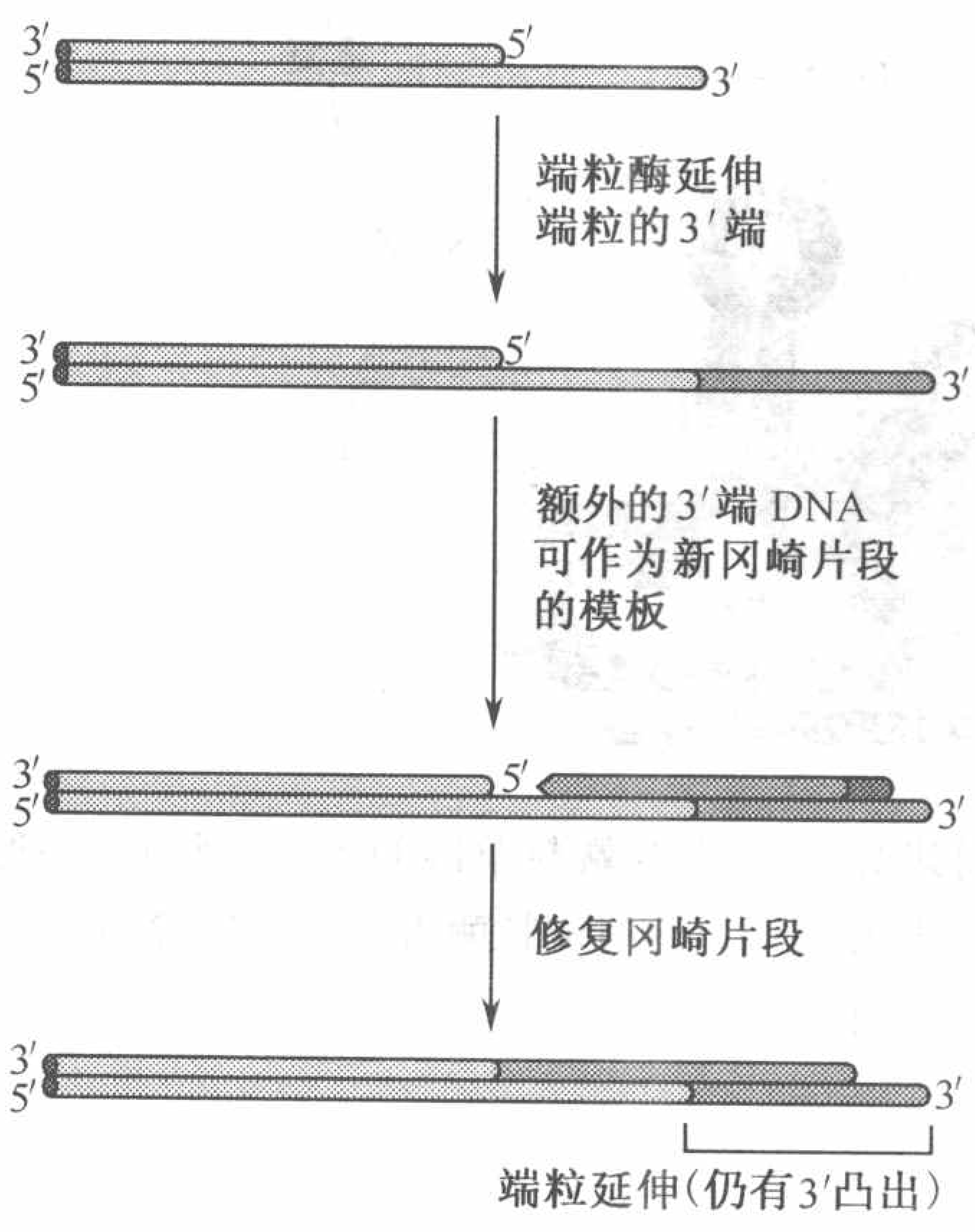


图 8-38 端粒酶对端粒 3′端的延伸解决了末端复制问题。虽然端粒酶只能通过为后随链 DNA 合成提供额外的模板来直接延伸端粒的 3′端，但这样染色体的两个末端都得到了延伸。

端粒结合蛋白调节端粒酶活性和端粒长度

理论上可以无限制地利用端粒酶延长端粒，结合于端粒双链区的蛋白质调节端粒的长度（图 8-39）。这些蛋白质或者结合在其上的其他蛋白质是端粒酶活性的微弱抑制剂（图 8-40）。当端粒序列重复拷贝较少时，这些蛋白质中的一部分与端粒结合并且端粒酶延长端粒的 3′羟基末端。随着端粒的逐渐增长，更多的端粒结合蛋白积累并且端粒酶延长端粒的 3′羟基末端。这一简单的负反馈机制（长端粒抑制端粒酶）是维持端粒在所有染色体末端相同长度的强有力的方法。

识别单链形式的端粒蛋白质同样能调节端粒酶活性。在酵母细胞中，Cdcl3 蛋白与端粒的单链区结合。研究表明该蛋白招募端粒酶到端粒上。因此 Cdcl3 是端粒酶的正向激活子。相比较而言，结合到单链端粒 DNA 上的蛋白质，POT1，则是端粒酶活性的抑制子。体外研究表明 POT1 结合到单链端粒 DNA 上抑制端粒酶的活性。缺少该蛋白的细胞端粒 DNA 的长度大大增加。有趣的是，该蛋白间接地和人细胞中双链端粒结合蛋白相互作用。研究已经表明端粒长度越长，更多的 TOP1 蛋白结合在端粒的单链 DNA 末端抑制端粒酶活性。

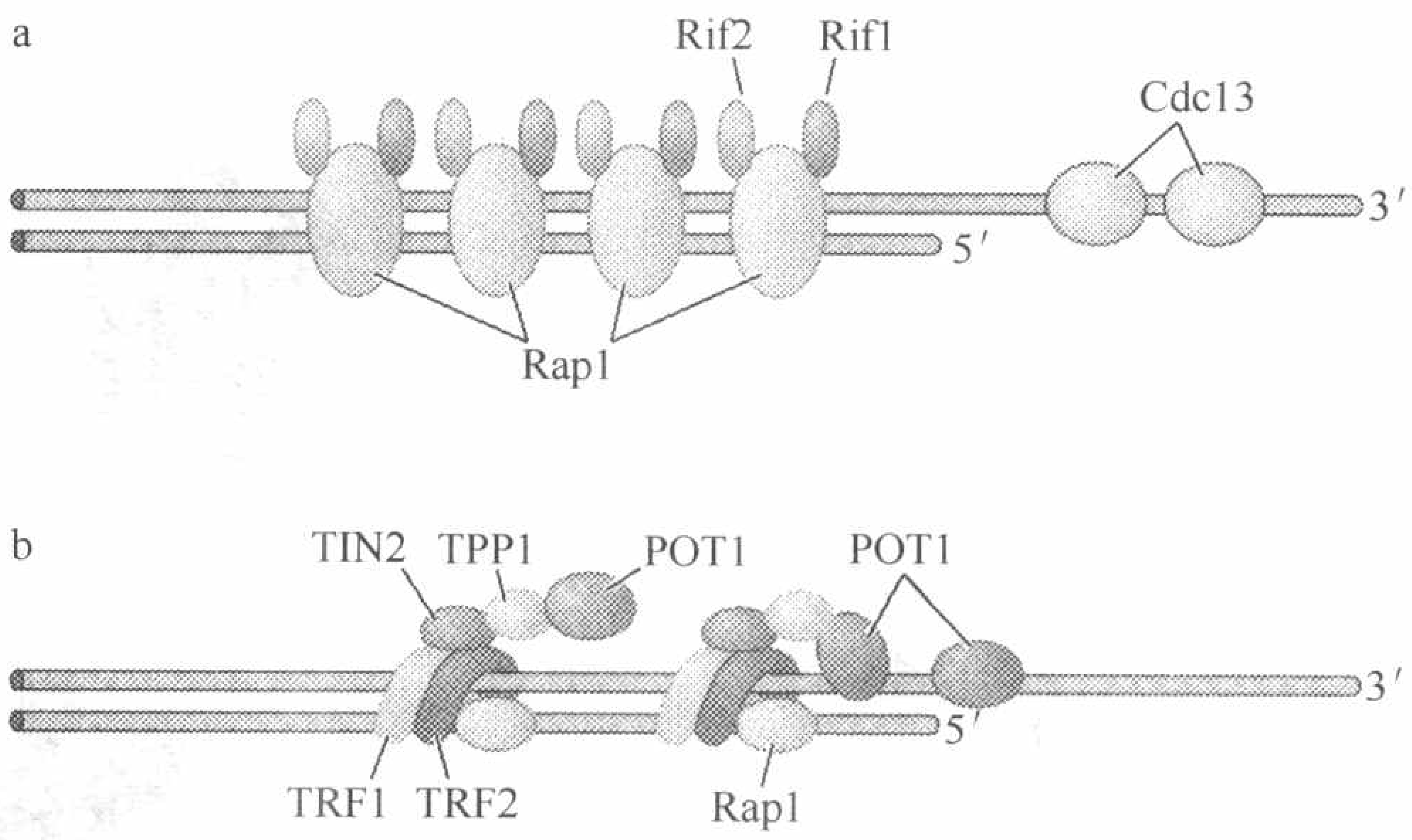


图 8-39　端粒结合蛋白。图示为酵母细胞和人细胞中调解端粒酶活性的端粒结合蛋白。（a）酵母细胞。Rap1 直接和双链端粒重复 DNA 结合，而 Rif1 和 Rif2 通过和 Rap1 的结合而间接地与端粒相互作用。这三种蛋白质已经用于研究端粒酶活性的抑制。Cdcl3 和单链端粒重复序列结合，在端粒酶的招募过程中起作用。（b）人细胞。TRF1 和 TRF2 直接结合到双链端粒重复 DNA 上。人细胞中 Rap1、TIN2、TPP1 和 POT1 的同源体都与 TRF1 和 TRF2 相互作用。这些蛋白质一起形成一个复合体，由于其对端粒酶的保护使端粒酶逃脱 DNA 修复酶的作用，此复合体被命名为“Shelterin”。POT1 同样能直接和单链端粒重复 DNA 结合。

端粒结合蛋白保护染色体末端

端粒结合蛋白除调节端粒酶功能之外，还对保护染色体末端至关重要。通常认为在细胞中，DNA 末端的存在是 DNA 中双链断裂的标记，该断裂是 DNA 修复机器的靶位点（见第 9 章）。这种修复最常见的结果是启动基因组上其他 DNA 之间的重组（在双倍染色体细胞中，这种重组以断裂染色体的完整拷贝为靶位点）。但是这种重组只适于

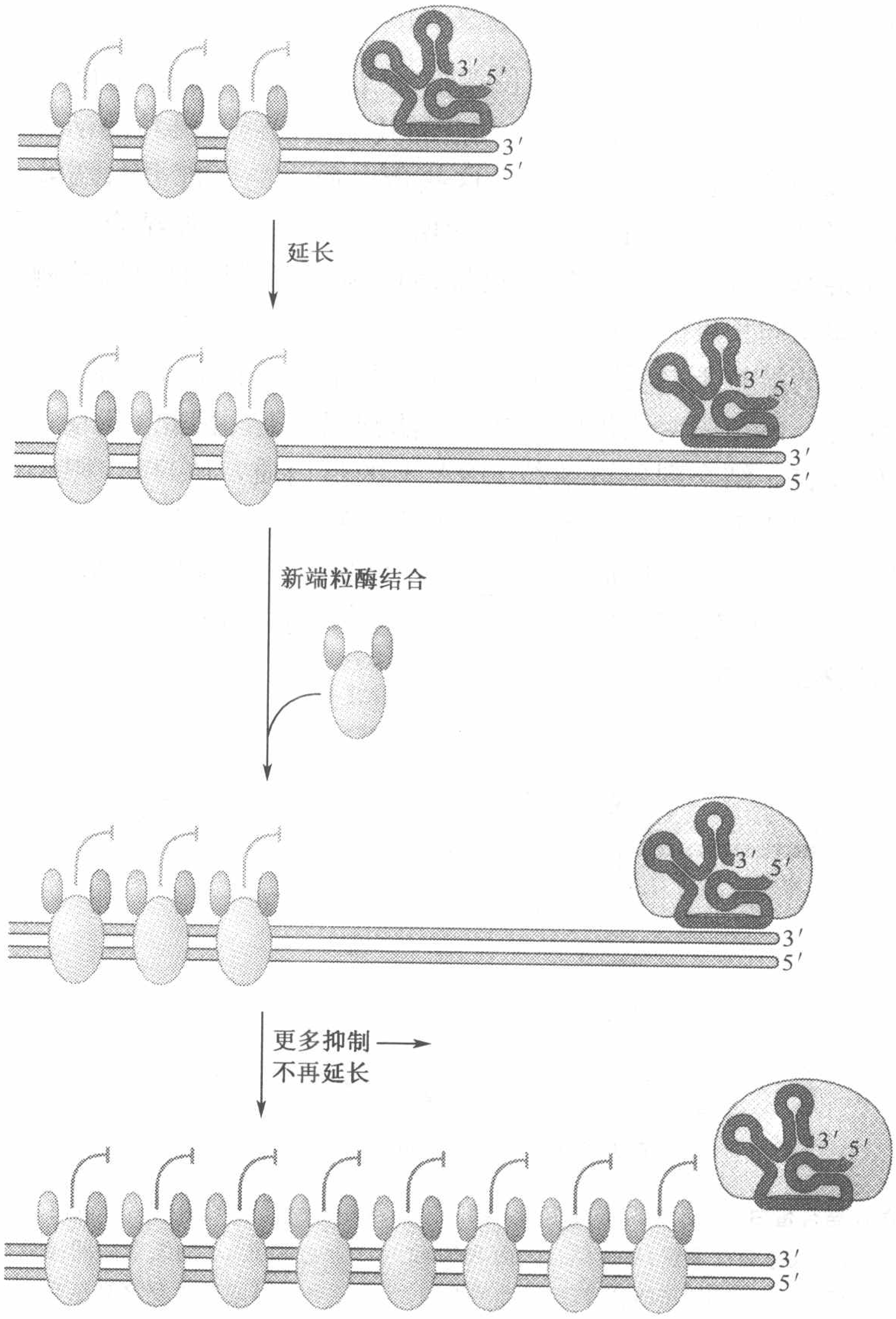


图 8-40 端粒结合蛋白调节端粒的长度。端粒酶相对较短时，少量端粒结合蛋白出现并对端粒酶有较弱的抑制作用。在此情况下，端粒酶可以延长端粒的 3′末端。当结合区域由后随链 DNA 合成机制作用变成双链结构时，其他端粒结合蛋白与端粒相互作用。该过程增强了对端粒酶抑制水平阻止端粒酶的进一步延伸。(经允许改编，源自 Smogorzeska A. and de Lange T 2004. Annu. Rec. Biochem. 73：177-208. Fig. 3a© Annual Reviews)

DNA 随机断裂，如果端粒参与该重组事件将会产生严重后果。以修复双链 DNA 断裂的方式来修复端粒会导致染色体融合，最终导致随机染色体断裂。

是什么在保护端粒逃脱这种命运呢？简单来说，结合在端粒上的蛋白质将端粒同细胞中其他 DNA 末端区分开来。这些蛋白质的减少造成端粒被误认为是正常的 DNA 断裂。一种简单可能的保护方法是用结合在其上的蛋白质包被端粒。对人类端粒结构的研

究提供了另一种保护方法。电镜观察从人类细胞中分离的端粒是环状结构而非线性结构（图 8-41a）。后续分析显示该结构（称为 t-loop）是由端粒的 3′单链 DNA 末端向端粒的双链 DNA 区域插入产生的。有趣的是，纯化的 TRF2 能够指导纯化的端粒 DNA t-loop 结构的形成。

t-loop 结构与端粒长度的控制有关。正如环状结构可以保护端粒不被 DNA 修复酶修复一样，端粒酶不能识别环状端粒，因为其缺少明显的单链 3′末端。有人提出端粒长度越短，端粒就越难形成 t-loop 结构，因此细胞拥有更多的端粒 3′末端。

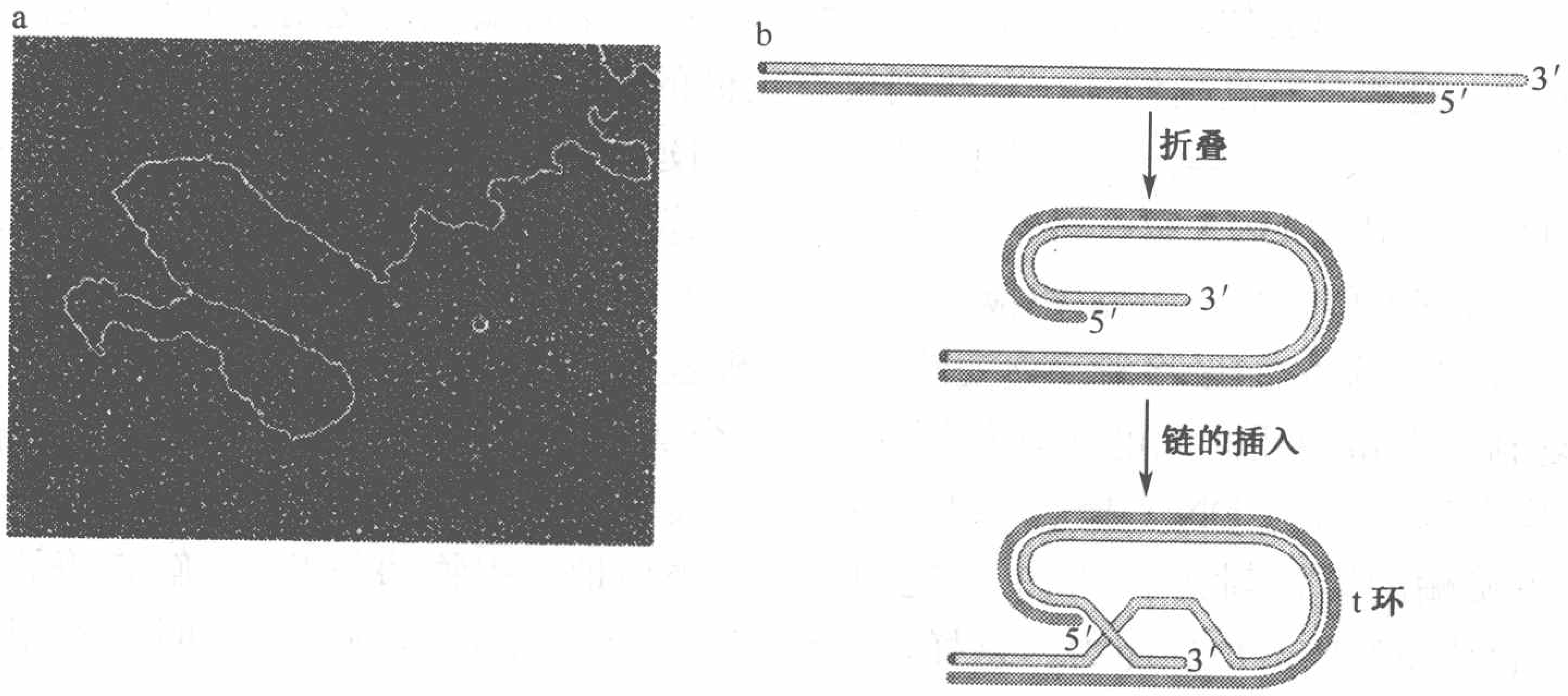


图 8-41　细胞中端粒环状结构的形成。（a）人类细胞中分离到的端粒的电镜图。DNA 末端环包括端粒末端的单链 DNA，即文中提到的 t-loop。图中右上角的 DNA 末端与染色体剩余部分结合。（b）t-loop 形成机制。首先端粒折叠使端粒末端的单链 DNA 接近端粒的双链 DNA 重复区。一旦单链 DNA 恰当定位，其侵入双链 DNA 重复序列，代替双链 DNA 另一条链与互补链形成螺旋。该现象称作置换环，是同源重组中常见的中间体（见第 10 章）。端粒末端蛋白和其他细胞蛋白（如重组蛋白）共同作用促进该过程的发生。注意端粒变短，此折叠更不易发生。（a，经允许再版，源自 Griffith J. D. et al. 1999. Cell 97：503-514. Fig. 3f © Elsevier）

小结

DNA 的合成依赖于两种类型的底物：4 种脱氧核苷三磷酸，dATP、dTTP、dCTP、dGTP；模板 DNA 的结构——引物-模板接头。模板 DNA 决定了掺入核苷酸的序列。引物作为脱氧核苷酸添加的底物，每个脱氧核苷酸都持续地添加到 3′端的羟基上。

DNA 的合成被称为 DNA 聚合酶的酶所催化，此酶利用单一活性位点来添加 4 种 dNTP 前体。对 DNA 聚合酶的结构研究发现其类似于一只抓住催化位点的手。此结构对极精确的 DNA 合成反应具有重要意义。DNA 聚合酶具有延伸性，每次结合一个底物，都可添加很多核苷酸。校正外切核酸酶通过类似于“删除键”那样的功能除去不正确添加的核苷酸，可进一步增强 DNA 合成的精确性。

在细胞中，DNA 模板的两条链在称为复制叉的结构上同时被复制。因为 DNA 的

两条链是反向平行的，所以只有一条 DNA 模板链能以连续方式进行复制（称为前导链）。另一条 DNA 链（称为后随链）必须先合成一系列短的称为冈崎片段的 DNA 片段。每条 DNA 链都用一条由引物酶所合成的 RNA 引物来起始。为完成复制，这些引物必须被除去。当 RNA 引物被 DNA 取代以后，所有单独合成的后随链 DNA 片段彼此连接在一起，形成一个连续的 DNA 链。

除 DNA 聚合酶以外，还有一系列的蛋白质帮助协调和促进 DNA 复制反应。这些额外的因子促进双链 DNA 模板的解旋（DNA 解旋酶）、稳定单链 DNA 模板（SSB），以及除去在复制叉前形成的超螺旋（拓扑异构酶）。DNA 聚合酶被特化用于 DNA 复制期间的不同事件。有些具有高延伸性，有些则仅有很低的延伸性。DNA 滑动夹增强了用于大范围 DNA 复制（如整条染色体）的 DNA 聚合酶的延伸能力。这些夹子蛋白与 DNA 拓扑连接，但是，当其与 DNA 聚合酶结合的时候能沿着刚合成出来的 DNA 滑行，有效地防止了结合的 DNA 聚合酶从引物-模板接头上脱落。称为滑行 DNA 夹子装载器的特殊蛋白复合体用 ATP 水解的能量将滑动夹安装到引物-模板接头附近的 DNA 上。

复制叉上蛋白质之间的相互作用在 DNA 合成中起着重要作用。在 *E. coli* 中，两个 DNA 聚合酶是称为 DNA 聚合酶Ⅲ全酶的大复合体的组分。DNA 聚合酶Ⅲ全酶与 DNA 解旋酶的结合刺激了 DNA 解旋的速度。类似地，引物酶与 DNA 解旋酶的结合增加了其合成 RNA 引物的能力。这样，当整套复制蛋白都在复制叉上的时候复制反应的效率最高。这套蛋白质所形成的复合体称为复制体。

DNA 合成的起始受复制器的特殊序列指导。复制起始的物理位点称为复制起始位点。复制器与起始子的蛋白质特异性地结合，起始子刺激起始位点处 DNA 的解旋并募集复制起始所需的其他蛋白质（如 DNA 解旋酶）。DNA 复制起始的过程主要由蛋白质-蛋白质或非特异性蛋白质-DNA 相互作用驱动。

真核细胞中 DNA 复制的起始被严格地调控，以确保每条染色体上的每个核苷酸在一轮细胞分裂中都只能被复制一次。这种严格的调控是通过控制前复制复合体（pre-RC）的多蛋白组合体的形成和激活来实现的。在复制器上形成这些复合体需要募集其他启动 DNA 复制所必需的蛋白质。形成和激活 pre-RC 的能力由一种称作细胞周期蛋白依赖性激酶的细胞周期调控激酶所控制。在细胞周期的 G_1 期，pre-RC 可形成但是不能指导复制的起始。在细胞周期的其他期内（S、G_2 和 M 期），任何已有的 pre-RC 都能启动 DNA 复制，但是不能形成新的 pre-RC。因此，一个细胞周期中，任何 pre-RC 都只能指导一轮起始，这保证了 DNA 只被精确地复制一次。

结束 DNA 复制需要特殊酶的作用。对于环形染色体来说，DNA 拓扑异构酶Ⅱ将拓扑连接的环状产物彼此分开；线性染色体也需要特殊的蛋白质来确保其复制的完整。在真核细胞中，一种称作端粒酶的特殊 DNA 聚合酶可使染色体末端（端粒）作为一个独特的复制起始位点。通过延长端粒的 3′端，端粒酶消除了复制叉机器进行常规合成时所造成的染色体末端渐进性丢失的问题。结合于端粒 DNA 上的蛋白调节端粒酶的活性，保护染色体破端不被降解和重组。

（张明　译　胡松年　刘韧　校）

参考文献

书

DePamphilis M.L. 2006. *DNA replication and human disease.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

DNA合成的化学基础

Brautigam C.A. and Steitz T.A. 1998. Structural and functional insights provided by crystal structures of DNA polymerases. *Curr. Opin. Struct. Biol.* **8:** 54–63.

Jäger J. and Pata J.D. 1999. Getting a grip: Polymerases and their substrate complexes. *Curr. Opin. Struct. Biol.* **9:** 21–28.

DNA聚合酶的作用机制

Doublié S. and Ellenberger T. 1998. The mechanism of action of T7 DNA polymerase. *Curr. Opin. Struct. Biol.* **8:** 704–712.

Steitz T.A. 1998. A mechanism for all polymerases. *Nature* **391:** 231–232.

———. 2006. Visualizing polynucleotide polymerase machines at work. *EMBO J.* **25:** 3458–3468.

复制叉

Corn J.E. and Berger J.M. 2006. Regulation of bacterial priming and daughter strand synthesis through helicase-primase interactions. *Nucleic Acids Res.* **34:** 4082–4088.

Johnson A. and O'Donnell M. 2005. Cellular DNA replicases: Components and dynamics at the replication fork. *Annu. Rev. Biochem.* **74:** 283–315.

O'Donnell M. and Kuriyan J. 2006. Clamp loaders and replication initiation. *Curr. Opin. Struct. Biol.* **16:** 405–415.

Wang J.C. 2002. Cellular roles of DNA topoisomerases. *Nat. Rev. Mol. Cell Biol.* **3:** 430–440.

DNA聚合酶的特化

Lemon K.P. and Grossman A.D. 2000. Movement of replicating DNA through a stationary replisome. *Mol. Cell* **6:** 1321–1330.

Lovett S.T. 2007. Polymerase switching in DNA replication. *Mol. Cell* **27:** 523–526.

Sutton M.D. and Walker G.C. 2001. Managing DNA polymerases: Coordinating DNA replication, DNA repair, and DNA recombination. *Proc. Natl. Acad. Sci.* **98:** 8342–8349.

DNA复制的起始

Bell S.P. and Dutta A. 2002. DNA replication in eukaryotic cells. *Annu. Rev. Biochem.* **71:** 333–374.

Blow J.J. and Dutta A. 2005. Preventing re-replication of chromosomal DNA. *Nat. Rev. Mol. Cell Biol.* **6:** 476–486.

Robinson N.P. and Bell S.D. 2005. Origins of DNA replication in the three domains of life. *FEBS J.* **272:** 3757–3766.

结束复制

Blackburn E.H., Greider C.W., and Szostak J.W. 2006. Telo-meres and telomerase: The path from maize, *Tetrahymena* and yeast to human cancer and aging. *Nat. Med.* **12:** 1133–1138.

Collins K. 2006. The biogenesis and regulation of telomerase holoenzymes. *Nat. Rev. Mol. Cell Biol.* **7:** 484–494.

第9章 DNA 的突变和修复

本章概要

•

复制错误及其修复

•

DNA 损伤

•

DNA 损伤的修复

遗传物质保持代代持续传递依赖于把突变概率维持在低水平上。生殖细胞系中高突变率将摧毁物种，体细胞中高突变率将摧毁个体。活细胞需要成千上万个基因正确行使职能，每个基因都可能毁于发生在其蛋白质编码序列或者负责其表达或信使 RNA 加工的旁侧序列等众多位点上的某个突变。

要使后代有很好的存活机会，DNA 序列必须在生殖细胞系中基本不变地进行传递。同样，如果生物成体中的体细胞中突变概率太高的话，这些细胞就不能执行其功能。例如，癌症是因为掌管细胞周期的基因受损，导致细胞无法按照可控方式进行生长和分裂而产生的。如果体细胞突变率高的话，癌症的发生率将是灾难性的和致命的。

同时，如果继承的遗传物质具有绝对的忠实性，将失去驱动进化所需的基因变异，那么新的物种，包括人类都不可能出现。因此，生命和生物多样性依赖于突变与突变修复之间的良好平衡。在本章中，我们将研究突变的原因以及负责逆转或纠正遗传物质上的这些损伤，从而使之最小化的系统。

突变的两个重要的来源是 DNA 复制的不准确性和遗传物质的化学损伤。来自互变异构化的复制错误我们已经在第 8 章见到，它为 DNA 复制过程中碱基配对的精确性设定了一个上限。复制 DNA 的酶促机器试图通过校正读码的机制来应对不正确核苷酸的错误插入，但是仍有些错误逃过检测。而且，DNA 是个复杂而且脆弱的有机分子，其化学稳定性有限。它不仅要遭受自发的损伤如碱基缺失，而且还遭受天然和非天然化合物以及辐射的攻击，可使其骨架断裂并改变其碱基的化学结构。简言之，复制错误和来自环境对遗传物质的损伤是不可避免的。第三个重要的突变源是被称为转座子的 DNA 元件造成的一种插入。转座就其本身来说是一个重要的课题，我们将在第 11 章进行详细阐述。

复制错误和 DNA 损伤有两个后果：一是给 DNA 带来永久性的改变（突变），这可能改变基因的编码序列或者基因的调控序列；二是 DNA 的某些化学变化使得 DNA 不能再用作复制和转录的模板。通常突变效果只在发生了序列改变的细胞的后代显示，但是妨碍复制或者转录的损伤能立即在细胞功能和存活上起作用。

细胞受到的挑战是双重的。第一，它必须扫描基因组以检查合成中的错误以及DNA的损伤；第二，它必须修复损伤，如有可能的话以一种恢复原始DNA序列的方式进行。本章我们将讨论复制过程中产生的错误、源于自发的DNA损伤，以及化学物质和辐射造成的损伤。每种情况下我们都要考虑遗传物质的改变如何被检测到以及它是如何被正确修复的。这些问题中我们将要进行讨论的如下：DNA如何足够快速地进行修复以防止错误成为突变存在于遗传物质中？在修复复制错误的过程中细胞如何将母链与子链区分开来？当断裂或者严重的损伤致使原始序列不能再被读取的时候，细胞如何恢复正确的DNA序列？细胞如何应对阻碍复制的损伤？回答这些问题有赖于了解需要进行修复的错误或损伤的类型。

我们将从探讨复制过程中发生的错误以及它们是如何被修复的开始，然后探讨自发产生的或来自外界的攻击造成的各种各样的损伤，再讨论细胞修复损伤的多种修复机制。我们将看到使细胞能处理DNA各种各样损伤的多个交叠系统，从而了解生物在保持遗传物质忠实性方面所做的努力。

复制错误及其修复

突变的本质

突变几乎包括DNA序列上每一种可以想像得到的改变。最简单的突变是一种碱基变成另一种碱基，有两种类型：转换（transition），即嘧啶到嘧啶或嘌呤到嘌呤的替换，如T到C和A到G；颠换（transversion），即嘧啶到嘌呤或嘌呤到嘧啶的替换，如T到G或A以及A到C或T（图9-1）。另一种简单的突变是一个核苷酸或少量核苷酸的插入或缺失。改变单个核苷酸的突变被称为点突变（point mutations）。

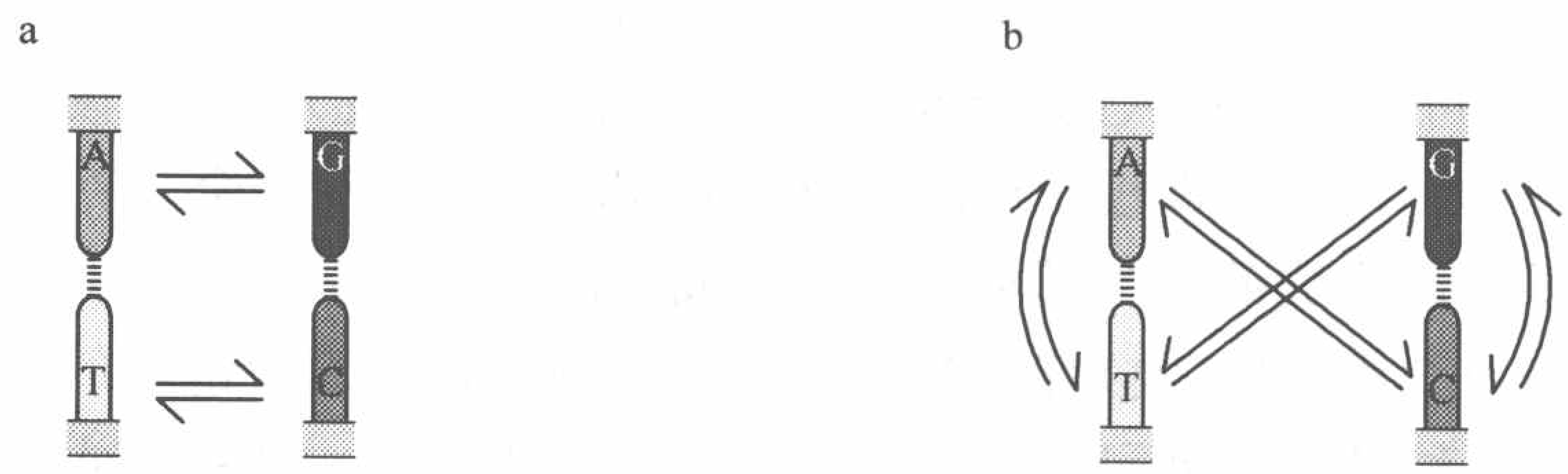


图9-1 碱基的替换。 a. 转换；b. 颠换。

其他类型的突变引起DNA的更大的改变，如大范围的插入和缺失以及染色体结构的整体重排。此种变化可能由转座子的插入或者细胞重组过程中的异常作用所导致——转座子通常将数千个核苷酸的外源DNA插入到基因的编码序列或者调控序列中（第11章）。染色体上任一给定位点自发产生新突变的总体概率为每轮DNA复制10^{-6}～10^{-11}，其中染色体上有些位点是“热点”，此处的突变发生频率高而其他位点发生变化

的概率相对较低。

有一种尤其易发突变的序列因为它对人类的遗传学和疾病的重要性而值得特别关注。这些突变易发序列是简单的 2-、3-或 4-核苷酸序列的重复，被称为 DNA 微卫星（DNA microsatellite）。一个为人熟知的例子涉及二核苷酸序列 CA 的重复。CA 重复片段在人类和其他真核细胞染色体上分布非常广泛。复制机器很难精确拷贝这种重复序列，常常发生“打滑”。打滑使得重复序列的拷贝数目增加或减少，结果，在人群中染色体上特定位置处的 CA 重复片段长度常常为高度多态性。这种多态性为可遗传突变作图（如增大人类患某些疾病倾向的突变）提供了方便的物理标记（框 9-1）。

框 9-1 三联核苷酸重复的扩增导致疾病

另一个为人熟知的错误易发序列的例子是某些基因中的三联核苷酸序列重复 CGG 和 CAG。常发现人类中的这种三联核苷酸重复在一代传给下一代的时候发生扩增，使得所导致的疾病在患病个体的孩子和孙子上渐进性加重。三联体扩增导致的疾病例子有成人肌（强直性肌）营养不良、导致精神迟缓的脆性 X 综合征，以及引起神经蜕变的亨廷顿病。CAG 是谷氨酰胺的密码子，它在亨廷顿蛋白编码序列中的扩增导致患有亨廷顿病的患者突变蛋白中含谷氨酰胺残基片段的长度加大。最近的研究表明，这个多聚谷氨酰胺片段干扰名为 Sp1 的转录因子中谷氨酰胺富含区与名为 TFIID 的转录机器元件的亚基“TAFII130”中对应的谷氨酰胺富含区之间正常的相互作用（见第 12 章）。此干扰削弱大脑神经元中的转录，包括神经递质受体基因的转录。其他基因中 CAG 扩增产生的相似的多聚谷氨酰胺片段可能也通过破坏转录因子与 TAFII130 之间的相互作用而作用。

有些复制错误能逃脱校正读码

如我们所知，复制机器凭借校正读码机制使精确度达到极高的水平，即复制子中的 3′→5′外切酶元件可将错误掺入的核苷酸去除（如第 8 章所述）。校正读码将 DNA 复制的忠实度提高约 100 倍，然而校正读码的核酸外切酶并非绝对安全。有些错误插入的核苷酸逃脱检测并在新合成的链与模板链之间形成错误配对。模板链中 4 种核苷酸每种的配对位置都可以有 3 种不同的核苷酸被错误插入（例如，在模板链上的 T 可能面对 T、G 或 C），共有 12 种可能的错配（T：T、T：G、T：C 等）。如果错误插入的核苷酸没有及时被检测到并被替换的话，序列的变化将永久性地存在于基因组中。在第二轮复制的时候，错误插入的核苷酸已经是模板链的一部分，将指导其互补性核苷酸插入到新合成的链中（图9-2）。此时，不存在错配，取而代之的是在 DNA 序列中产生一个永久性的改变（突变）。

错配修复能将逃脱校正读码的错误去除

幸运的是，有一套用来检测错配并对之进行修复的机制。保证 DNA 复制忠实性的

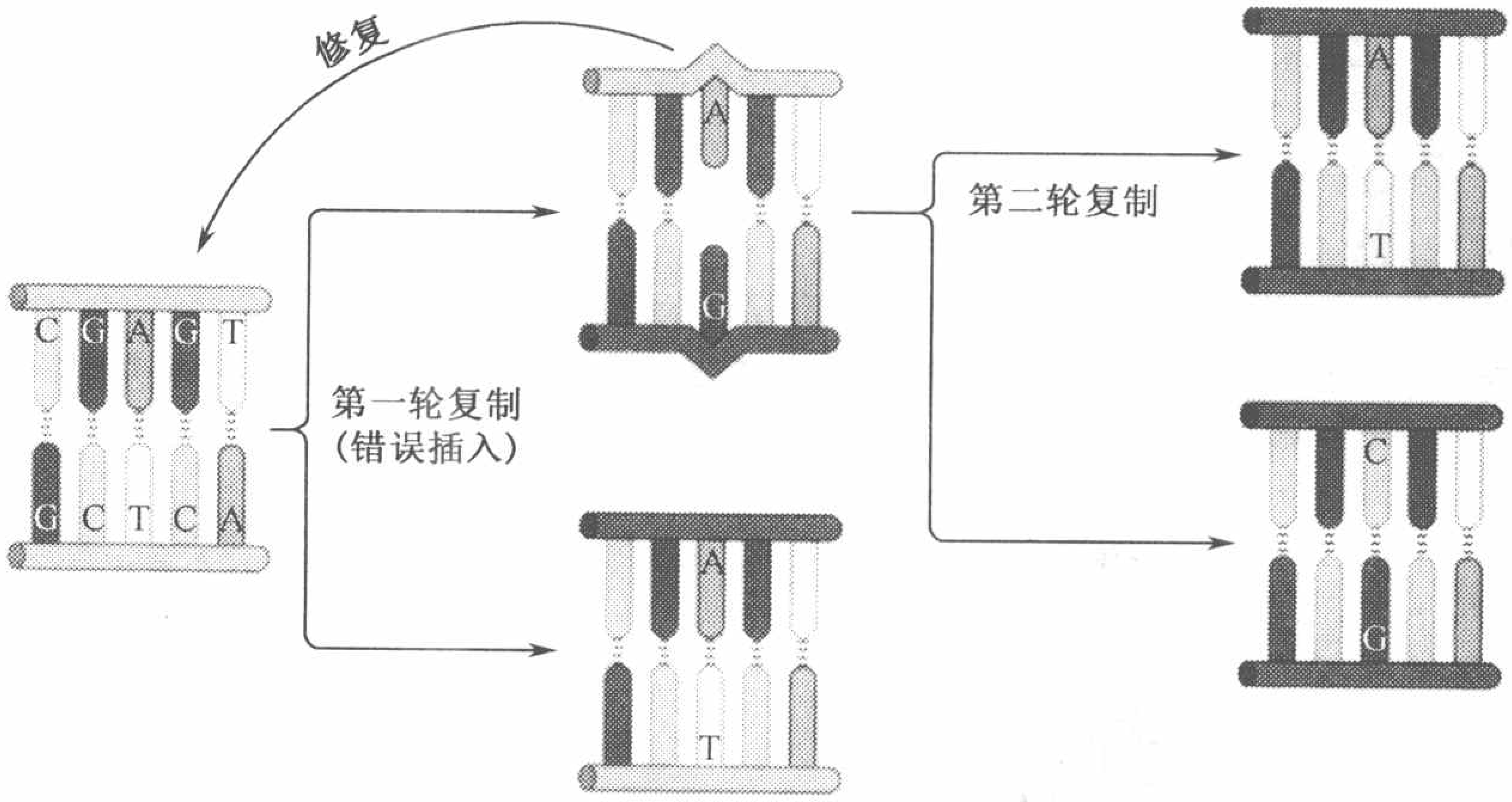


图 9-2 突变可通过复制永久性地插入进去。在第一轮复制中错误插入一个碱基可引入一个突变；在第二轮复制中，突变就永久性地插入到 DNA 序列中了。

最后责任由错配修复系统（mismatch repair system）承担，它将 DNA 合成精确性又提高了2～3个数量级。错配修复系统面对两个挑战：第一，它必须扫描基因组寻找错误配对，因为错误配对是短暂的（它们在第二轮复制后被消除，此时它们产生突变），错配修复系统必须很快就发现错误配对并对之进行修复；第二，系统必须准确地纠正错误配对，即它必须在新合成的链中将错误插入的核苷酸换掉而不能把母链中正确的核苷酸换掉。

在大肠杆菌中，错配是由错配修复蛋白 MutS 的二聚体来检测的（图 9-3；见结构指南 9-1）。MutS 扫描 DNA，从错配引起 DNA 骨架的扭曲变形上识别出它们。MutS 包围含错误配对的 DNA，诱导 DNA 产生一个明显的纽结和 MutS 自身构象上的变化（图 9-4）。MutS 特异性的关键在于含有错误配对的 DNA 比正确碱基配对的 DNA 更容易扭曲。MutS 拥有错配修复所需要的 ATP 酶活性但是它在修复中的具体作用还是未知的。MutS 和含错误配对的 DNA 形成的复合体募集 MutL——错配修复系统中的第二个蛋白质。MutL 再激活 MutH。MutH 是一种酶，能在错配位点附近的一条链上产生一个切口或割口。切口形成后由一种特殊的 DNA 解旋酶（UvrD）和三个核酸外切酶中的一个发挥作用（见下文）。解旋酶从切口开始向错配位点方向解开 DNA，核酸外切酶逐步地将解旋的单链消化掉，延伸到并越过错配核苷酸的位点。此作用产生一个单链缺口，此缺口被 DNA 聚合酶Ⅲ（Pol Ⅲ）填补并由 DNA 连接酶接合。整体效果是将错误配对除去并用正确碱基配对的核苷酸取而代之。

但是 *E. coli* 错配修复系统如何知道两个错配核苷酸中的哪一个要被取代呢？如果修复是随机发生的话，那么有一半的概率错误将永久存在于 DNA 中。答案是 *E. coli* 通过我们就要提到的暂时半甲基化对母链进行了标记。

E. coli 的 Dam 甲基化酶（DAM methylase）使两条链上的 5′-GATC-3′ 序列中的 A 残基都甲基化。GATC 序列广泛分布于整个基因组中［约每 256（4^4）个碱基对发生一

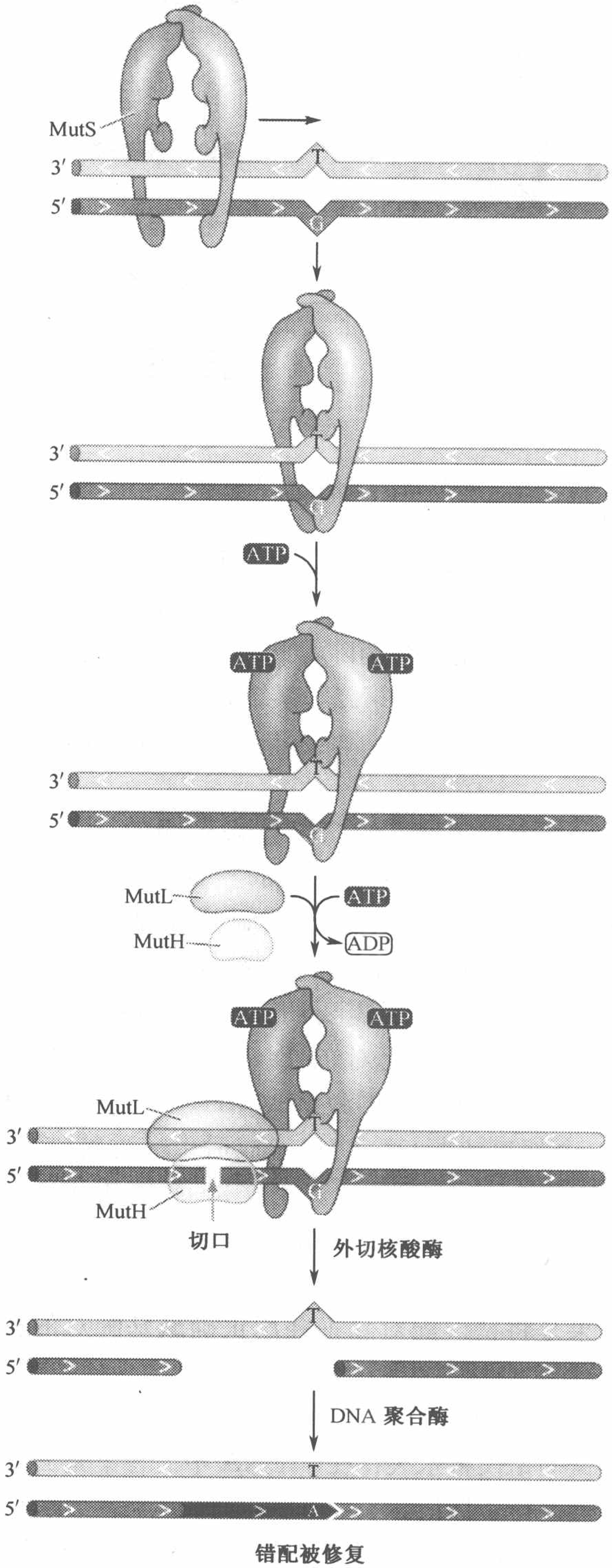


图 9-3 用以修复复制错误的差误配对修复途径。 MutS 包围着含有纽结的错配 DNA（未标出，见图 9-4）。接下来，MutS 募集 Mut L 和 MutH 并且由 MutS 的 ATP 酶活性催化 ATP 的水解。MutH 是一种内切核酸酶能在 DNA 上接近错配的位置产生一个切口。紧接着，一个外切核酸酶消化切口链，向着错配方向移动。最后，产生的单链的缺口由 DNA 聚合酶填补并消除错配。（经允许改编自 Junop M. S. et al. 2001. Mol. Cell 7：1-12，Fig. 6b. © Elsevier.）

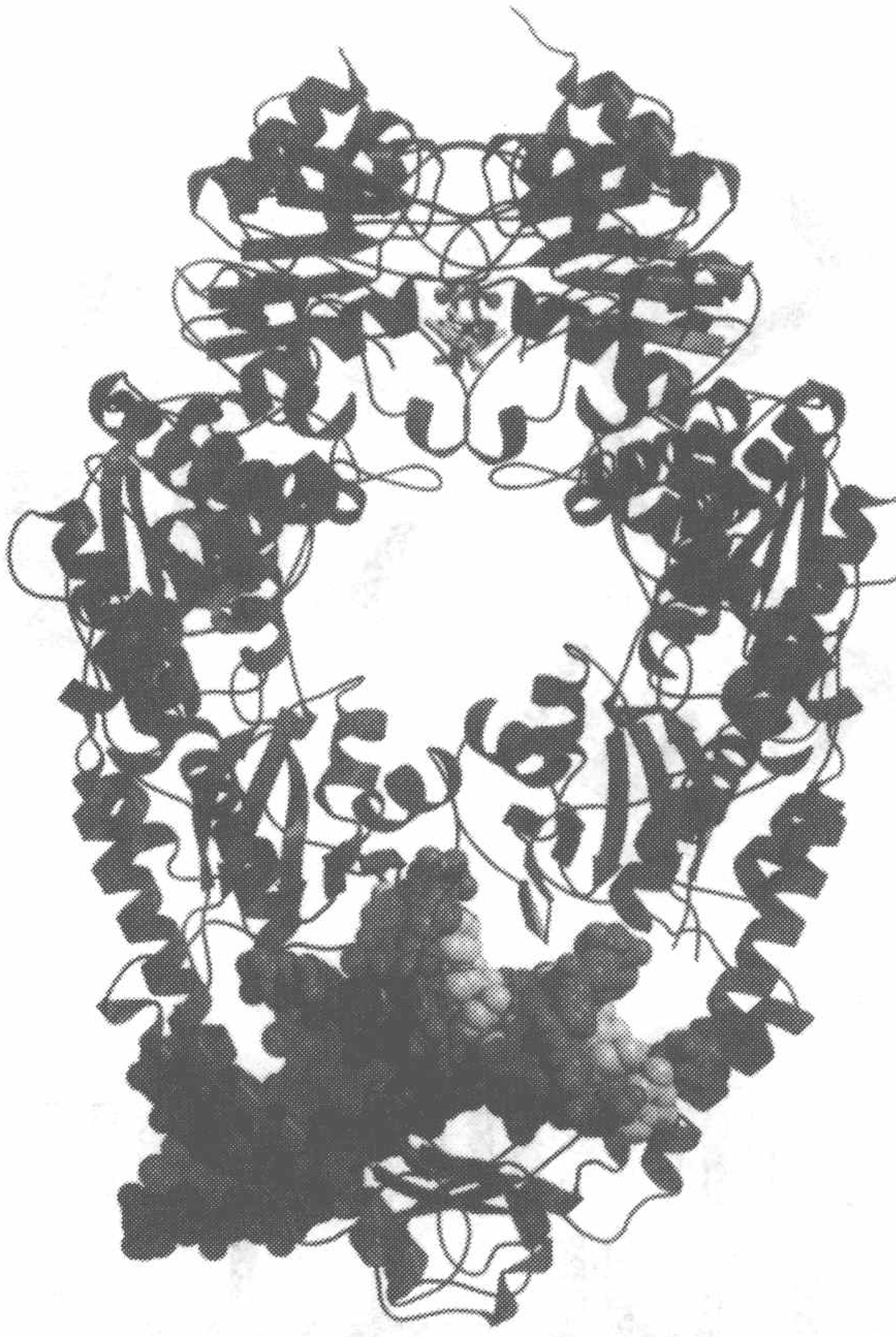


图 9-4 MutS-DNA 复合体的晶体结构。注意 DNA 上位于结构底部的纽结和在酶的结构的顶部的 ATP，分别用黄色、绿色和红色表示。DNA 用一种空间填充的形式表示，红色代表骨架，灰色代表碱基。（Junop M. S. et al. 2002. Mol. Cell 7：1-12.）镜像图用 BobScript、MolScript 和 Raster 3D 绘制。

次]，所有这种位点都被 Dam 甲基化酶甲基化。当复制叉经过两条链的 GATC 都甲基化（全甲基化）的 DNA 时，产生的子代 DNA 双链就是半甲基化的（只有母链是甲基化的）。这样在几分钟的时间内，直到 Dam 甲基化酶赶上并将新合成的子链甲基化为止，子代 DNA 双链将只有一条担任模板的链被甲基化（图 9-5a）。这样，新合成的链即被标记（它没有甲基基团）从而被认出是要被修复的链。

MutH 蛋白结合在这样的半甲基化位点处，但它的核酸内切酶活性通常是隐藏的。只有被位于附近错配处的 MutL 和 MulS 接触的时候（这可能是在几百个碱基对的距离内），MutH 才能像我们上面所说的那样被激活。这种相互作用如何能在数百 bp 的距离以外发生，还是未知的。但是最近有证据表明，MutS-MutL 复合体可以离开错配处并沿着 DNA 的外缘移动到半甲基化位点处的 MutH。MutH 被激活后，选择性地将未甲基化的链切开，这样在错误配对附近只有新合成的 DNA 被除去并被取代（图 9-5b）。因此甲基化是在复制发生错误的时候，使 *E. coli* 修复系统能够从母链上找回正确序列的“记忆”设备。

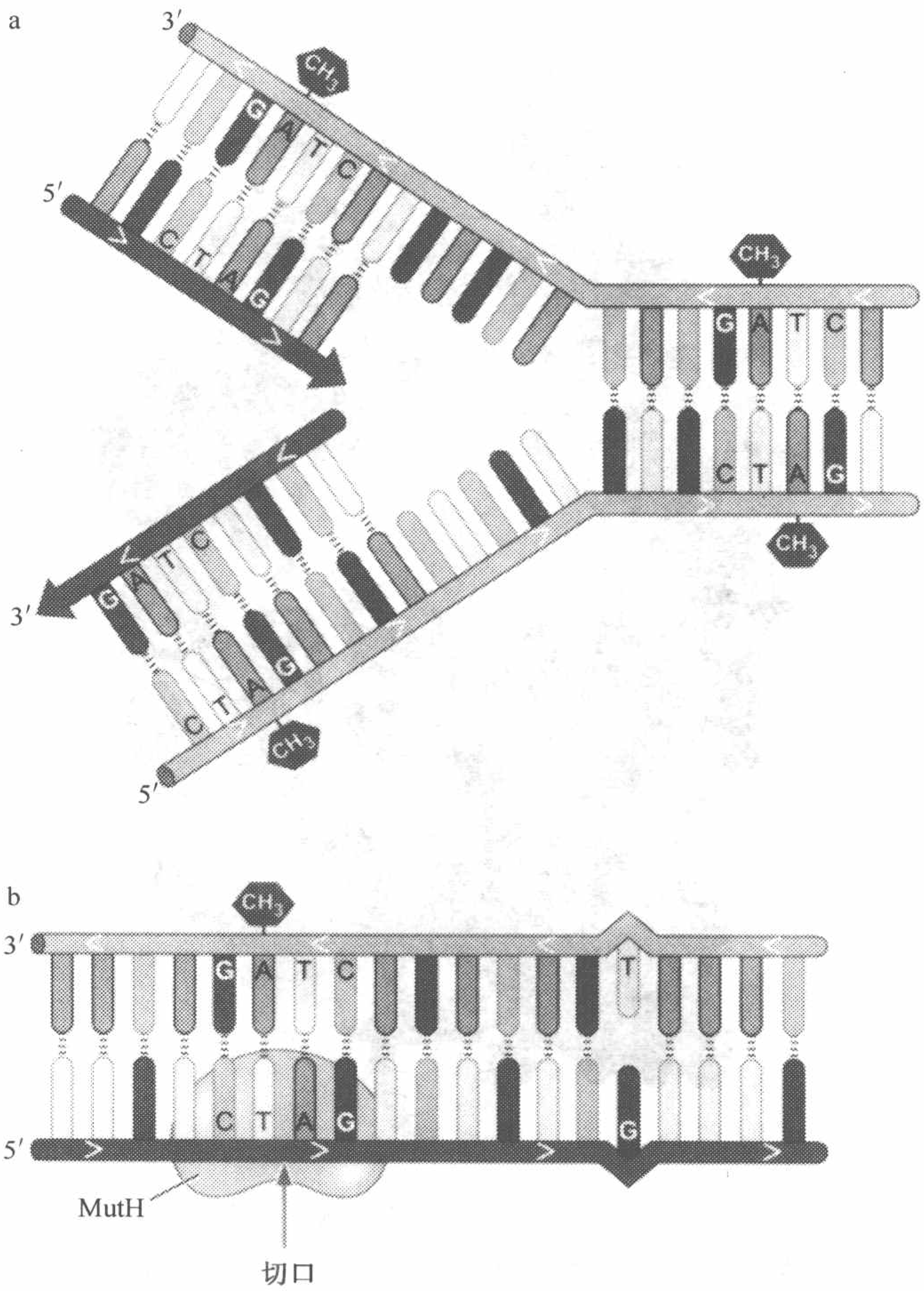


图 9-5 Dam 在复制叉上的甲基化。 a. 复制在 *E. coli* 中产生半甲基化的 DNA。b. MutH 在未甲基化的子链上进行切割。

MutH 产生的切口与错误配对之间单链 DNA 的去除可以使用不同的核酸外切酶，这取决于 MutH 在错误插入的核苷酸 5′还是 3′一侧切割 DNA。如果在错配的 5′侧切割 DNA，则用核酸外切酶Ⅶ或 RecJ 沿 5′→3′方向从 MutH 产生的切口开始经过错误插入的核苷酸来除去 DNA 片段。反之，如果切口在错配的 3′侧，则由核酸外切酶Ⅰ沿 3′→5′方向除去 DNA 片段。如我们所看到的，错配碱基被除去之后，由 DNA 聚合酶Ⅲ来填补缺失的序列（图 9-6）。

真核细胞也能修复错误配对，它们使用 MutS（MutS 同源物称为 MSH 蛋白）和 MutL（MutL 同源物称为 MLH 和 PMS）的同源物完成这一功能。事实上，真核生物有多种具有不同专一性的 MutS 样蛋白。例如，一种是专门用于简单错配的，而另一种用于识别 DNA 复制期间“打滑”产生的小的插入或缺失。错配修复在较高等生物上扮演极为重要的角色，最生动的证据是基于发现结肠癌（遗传性非息肉性结直肠癌）的遗传倾向性是由于人的 MutS 同源物（尤其是 MSH2 同源物）和 MutL 同源物的基因突变。

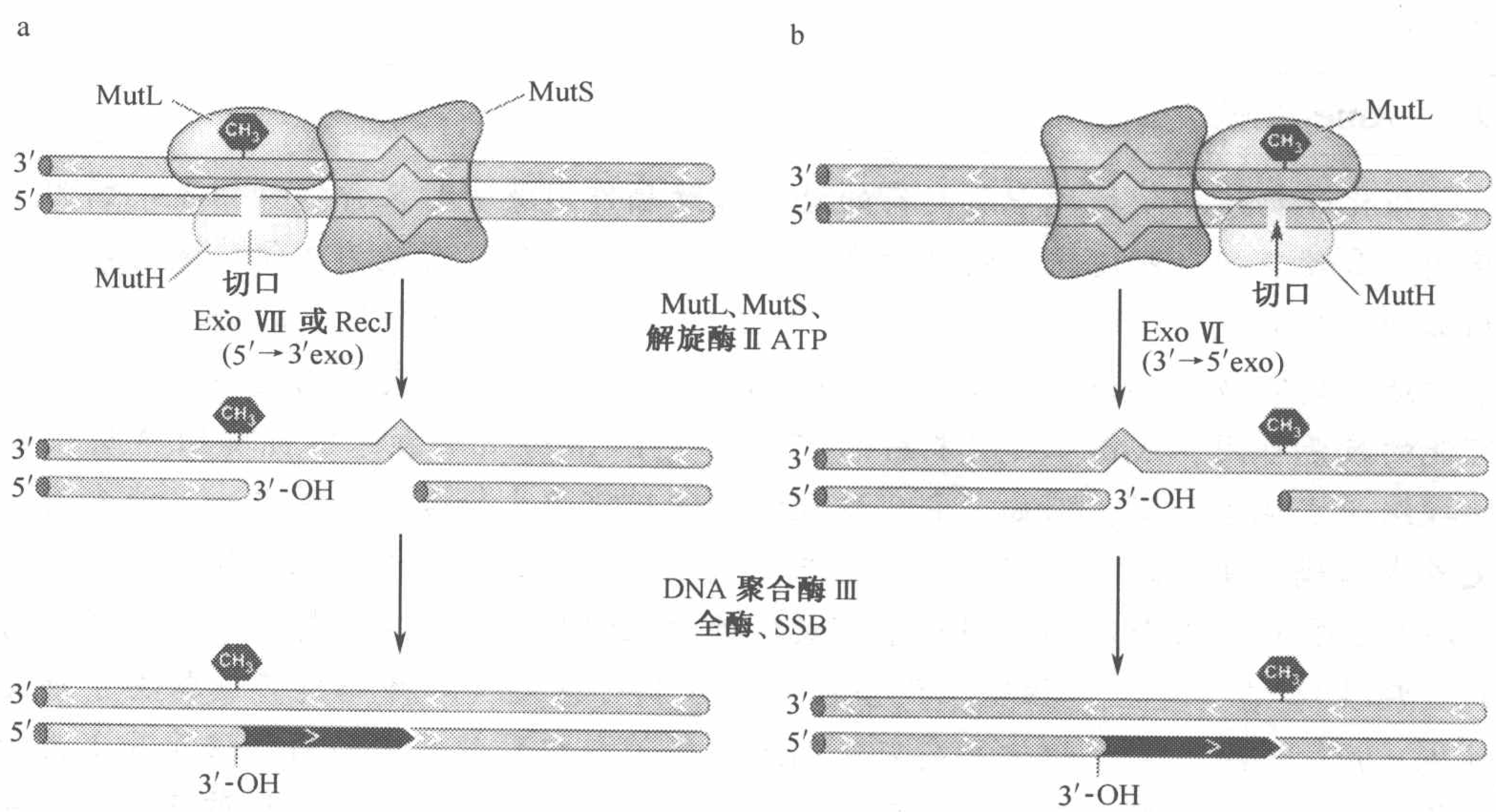


图 9-6 错配修复的方向性：核酸外切酶对错配 DNA 的去除。为了简化，与 DNA 结合的 MutH 表示成与错配处的 MutS 直接相连的形式。a. 未甲基化的 GATC 位于突变的 5′侧；b. 未甲基化的 GATC 位于突变的 3′侧。

即使真核生物细胞有错配修复系统，它们还缺少 MutH 以及像 *E. coli* 那样用半甲基化标记母链的聪明策略。(事实上，多数细菌都没有 Dam 甲基化酶，并且也不能用半甲基化来标记新合成的链。) 那么错配修复系统是怎么知道两条链中哪一条需要校正的呢？如我们在第 8 章中看到的那样，后随链进行不连续合成产生冈崎片段，再通过 DNA 连接酶与先前合成的 DNA 相连接。在连接步骤之前，冈崎片段与先前合成的链之间隔着一个切口，这个切口可以看作等同于 *E. coli* 中 MutH 在新合成链上产生的切口。事实上，真核细胞的提取物将修复含有切口的人工模板上的错配，并且是选择性地在携带切口的链上进行。最近的结果表明，MutS 的人类同源物（MSH）与复制体上的滑动夹元件(PCNA，我们在第 8 章曾予以讨论）相互作用，因此可能被招募到后随链 DNA 不连续合成的部位上。与滑动夹的相互作用也可能将错配修复蛋白招募到前导链的 3′（生长中）端。

DNA 损伤

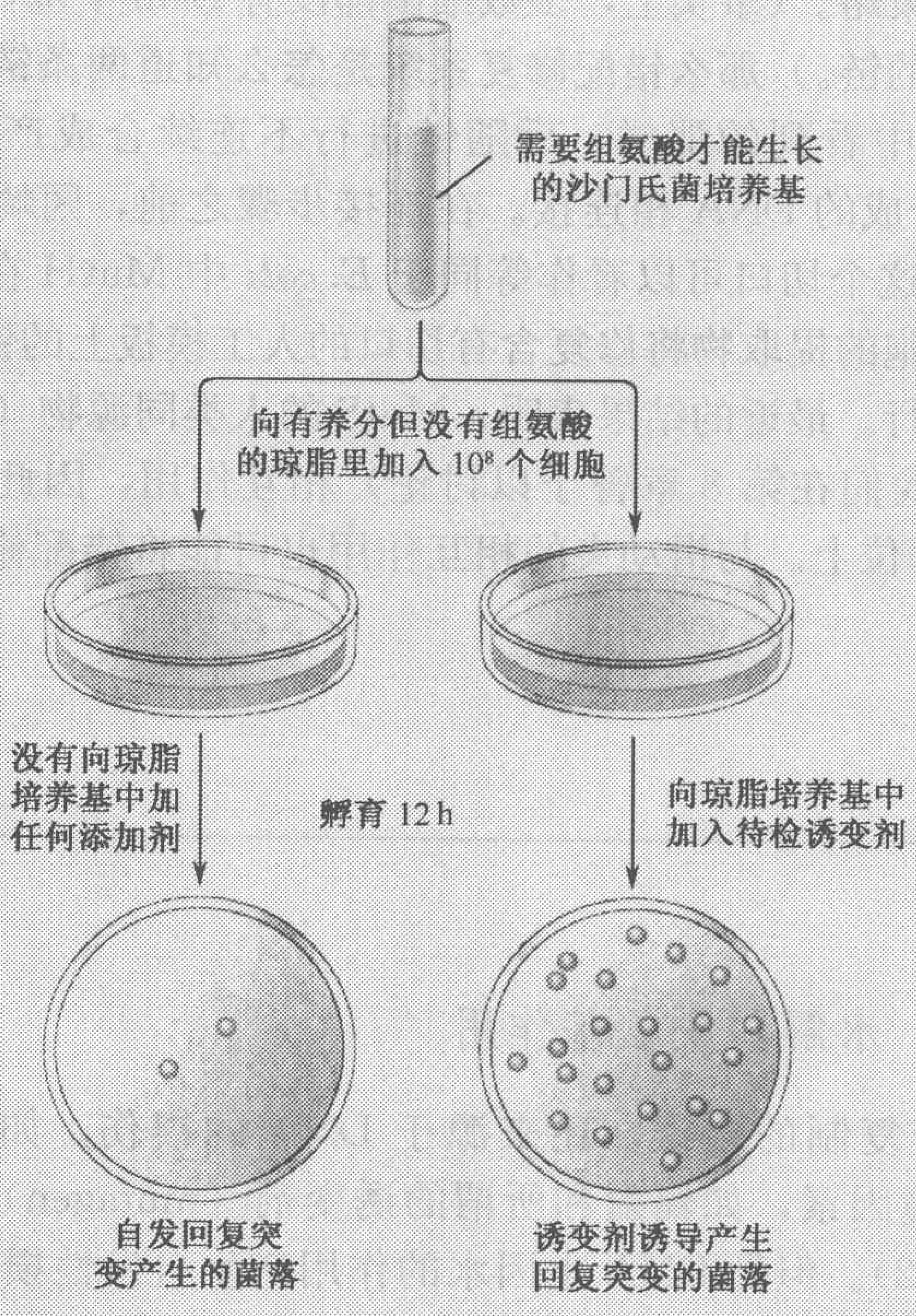
DNA 的自发损伤：水解和脱氨基作用

突变不仅来源于复制的错误，还来源于 DNA 的损伤。如我们将看到的，有些 DNA 损伤来源于环境因素，如辐射和所谓的诱变剂（mutagen）——增加突变发生概率的化学试剂（框 9-2)。但是 DNA 也因水的作用产生自发性损伤（这听起来很滑稽，因为双螺旋的正确结构有赖于水性的环境）。

医学联系

框 9-2 Ames 测试

分析化学物质在动物体中所具有的潜在的致癌作用既费时又费钱。但是，因为多数致癌剂都是致变剂，所以化学物质潜在的致癌作用可从其致变能力方便地进行评估。伯克利加利福尼亚大学（UCBerkeley）的 Bruce Ames 设计了一个简单的测试方法，根据化学物质在伤寒沙门氏菌（*Salmonella typhimurium*）中引起突变的能力来评价化学物质可能的致癌作用。Ames 测试采用的这个 *S. typhimurium* 菌株中负责组氨酸生物合成的操纵子发生了突变。例如，突变株操纵子的组氨酸生物合成基因之一可能含有一个错义或者移码突变。结果，突变株细胞在组氨酸缺乏的固体培养基上不能生长和形成菌落（框 9-2 图 1）。但是，如果用一个有致变性的化学物质处理突变株细胞（因而具有潜在的致癌性），在少量突变细胞中化合物将使原来的错义或者移码突变（取决于致变剂的性质）产生回复突变。此回复突变使细胞恢复了在组氨酸缺乏的固体培养基上生长和形成菌落的能力。诱变剂越强，菌落的数目就越多。有些致癌的化学物质开始的时候没有致变性，但是代谢外源物质的肝脏将其转变成了致变剂。为鉴别能够在肝脏中转化为突变剂的化学物质，Ames 中肝中各种酶的混合物处理潜在的致变剂。Ames 测试中发现有致变性的化学物质可进一步在动物身上再测试其潜在的致癌作用。



框 9-2 图 1 Ames 实验

最频繁和最重要类型的水解损伤是胞嘧啶碱基的脱氨基（图 9-7a）。在正常的生理条件下，胞嘧啶可自发进行脱氨基作用，从而产生非天然的（在 DNA 中）碱基尿嘧啶。尿嘧啶优先与腺嘌呤配对，因此，复制时在另一条链上导入腺嘌呤，而不是胞嘧啶指导下的鸟嘌呤。腺嘌呤和鸟嘌呤也易于自发脱氨基，脱氨基将腺嘌呤转变成次黄嘌呤，次黄嘌呤与胞嘧啶而不是与胸腺嘧啶形成氢键；鸟嘌呤转变成黄嘌呤，黄嘌呤虽然与胞嘧啶之间只有两个氢键，但是它仍与胞嘧啶配对。DNA 还通过 *N*-糖苷键的自发水解进行去嘌呤化（depurination），并在 DNA 中形成脱碱基（abasic）位点（没有碱基的即脱氧核糖）（图 9-7b）。

注意，与上面所讨论的复制错误不同，这些水解反应都导致 DNA 形成非天然的改变。无嘌呤位点当然是非天然的，而且每个脱氨基作用都产生一个非天然的碱基，因而这种改变可被下面所述的修复系统识别。这种情况还说明了为什么 DNA 有胸腺嘧啶而没有尿嘧啶。如果 DNA 天然含有尿嘧啶而不是胸腺嘧啶的话，那么胞嘧啶的脱氨基作用将产生一个天然碱基，这是修复系统很难识别的。

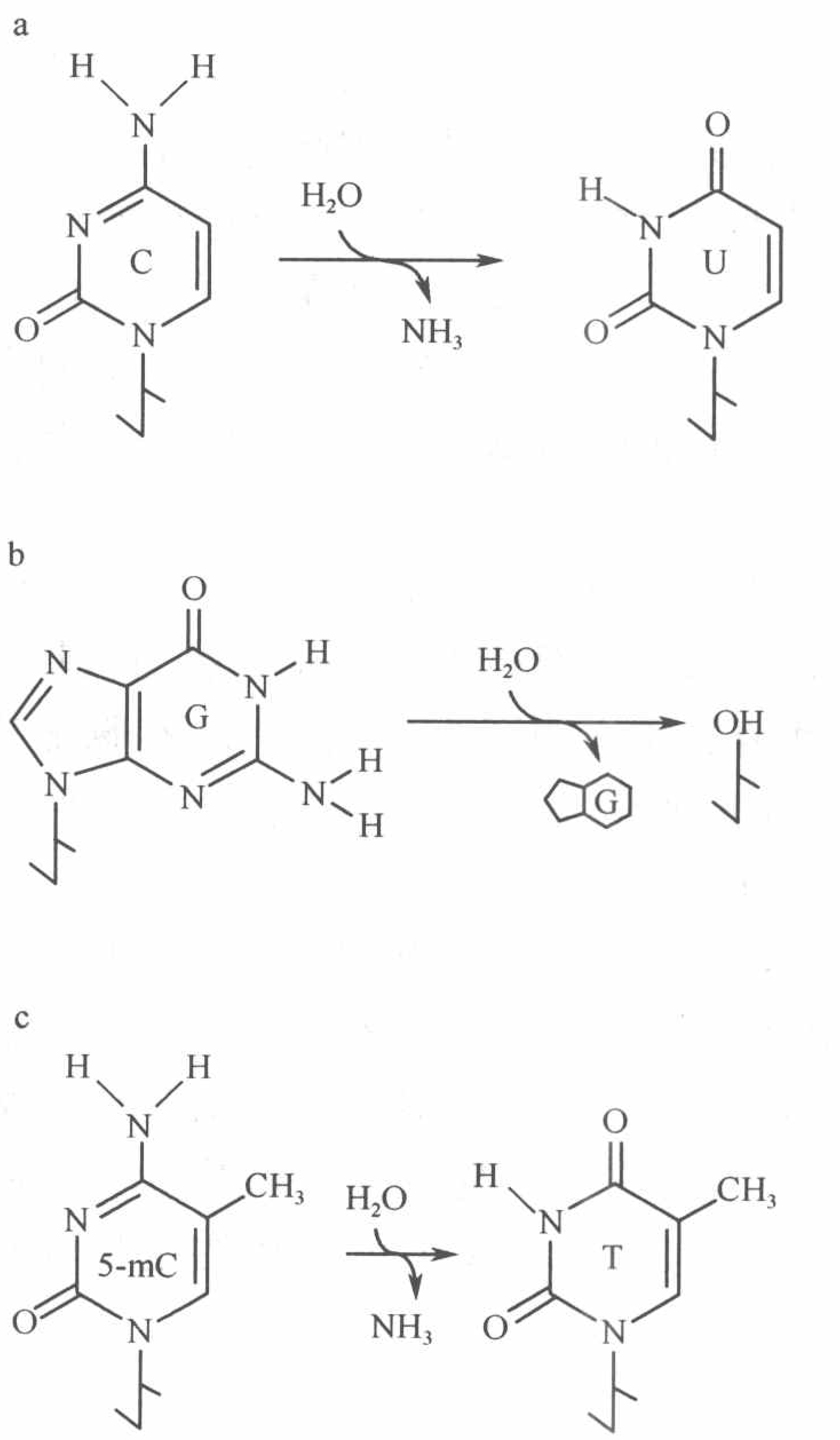
脱氨基作用产生一个天然存在碱基的危险，可用 5′-甲基胞嘧啶引起的问题来解释。脊椎动物 DNA 由于甲基转移酶的作用经常在胞嘧啶的位置上含有一个 5′-甲基胞嘧啶。这个修饰后的碱基在转录沉默中起作用（第 17 章）。5′-甲基胞嘧啶脱氨基产生胸腺嘧啶（图 9-7c），明显不能被识别为异常碱基，所以，在下一轮复制中就作为胞嘧啶到胸腺嘧啶（C→T）的转换而被固定。事实上，脊椎动物 DNA 中甲基化的胞嘧啶是自发突变的热点。

DNA 损伤：烷化反应、氧化反应和辐射

DNA 易于受烷化反应、氧化反应和辐射的损伤。在烷化反应中，甲基或乙基基团被转移到碱基的反应位点以及 DNA 骨架的磷酸上。烷化化合物包括亚硝胺（nitrosamines）和非常强烈的实验室诱变剂 *N*-甲基-N^1-硝基-*N*-亚硝基胍。（*N*-methyl-*N*′-nitro-*N*-nitrosoguanidine）鸟嘌呤 C6 原子上的氧是最容易发生烷基化的位点之一（图 9-8）。此甲基化的产物——O^6-甲基鸟嘌呤（O^{-6}-methyl-guanine）常常与胸腺嘧啶错配，导致受损 DNA 复制的时候 G∶C 碱基对变成 A∶T 碱基对。

DNA 还易受到活性氧簇的攻击（如 O_2^-、H_2O_2 和 OH·）。这些强烈的氧化剂由电离辐射和产生自由基的化学试剂产生。例如，鸟嘌呤氧化后产生 7，8-二羟基-8-氧代鸟嘌呤或简称氧代鸟嘌呤（oxoG）。氧代鸟嘌呤有强烈的致变性，因为它既能与腺嘌呤也能与胞嘧啶碱基配对。如果在复制的时候与腺嘌呤碱基配对，则产生 G∶C 到 T∶A 的颠换，这是人类癌症中最普通的突变之一。因此，电离辐射和氧化剂的致癌效应可能部分是由于能把鸟嘌呤转换成氧代鸟嘌呤的自由基引起的。

另一种碱基损伤是紫外线造成的。260nm 左右波长的辐射被碱基强烈吸收，其后果之一是在同一多核苷酸链相邻位置上的两个嘧啶之间发生光化学聚合。两个胸腺嘧啶的聚合产物叫做胸腺嘧啶二聚体（thymine dimer）（图 9-9），含有由毗邻胸腺嘧啶的 C5 和 C6 之间形成的环丁烷（cyclobutane）环。当胸腺嘧啶与胞嘧啶毗邻时，产生的聚合体是胸腺嘧啶-胞嘧啶加合物，此处胸腺嘧啶通过其 C6 连接到胞嘧啶的 C4 上。



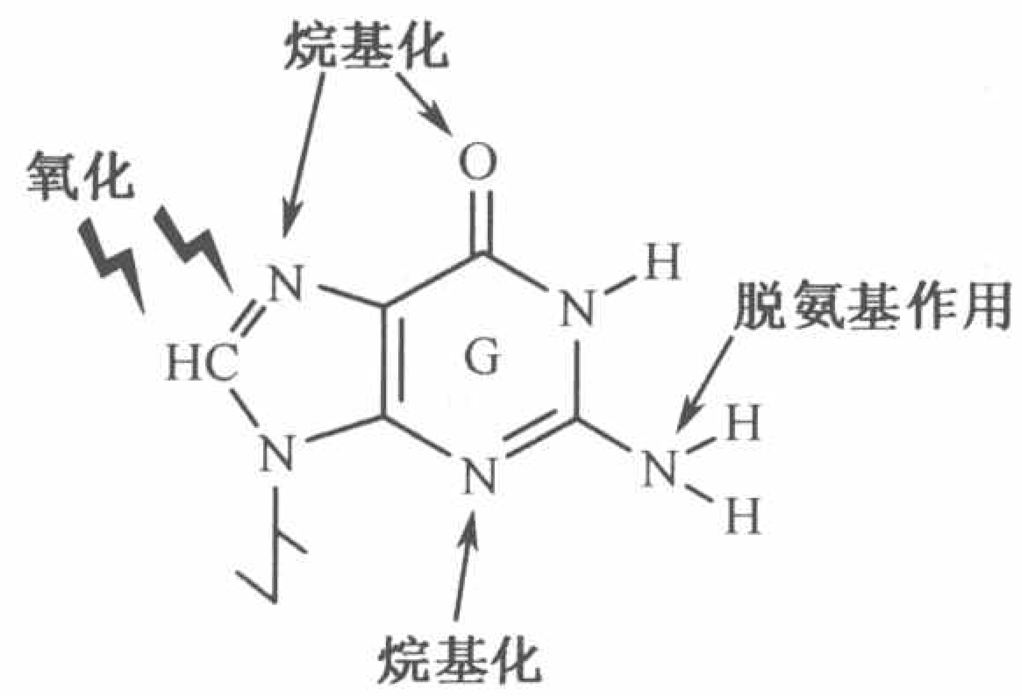


图 9-7 水解损伤引起的突变。 a. 胞嘧啶脱氨基产生鸟嘌呤。b. 鸟嘌呤通过水解脱嘌呤化产生脱嘌呤的脱氧核糖。c. 5-甲基胞嘧啶的脱氨基在 DNA 中产生天然碱基——胸腺嘧啶。

图 9-8 鸟嘌呤的修饰。 图示为鸟嘌呤上易于被化学处理，如烷化反应或氧化反应以及辐射损伤的特殊部位。经这些修饰的产物常常为高度致变的。

这些聚合的碱基不能再进行碱基配对并导致复制过程中 DNA 聚合酶的终止。

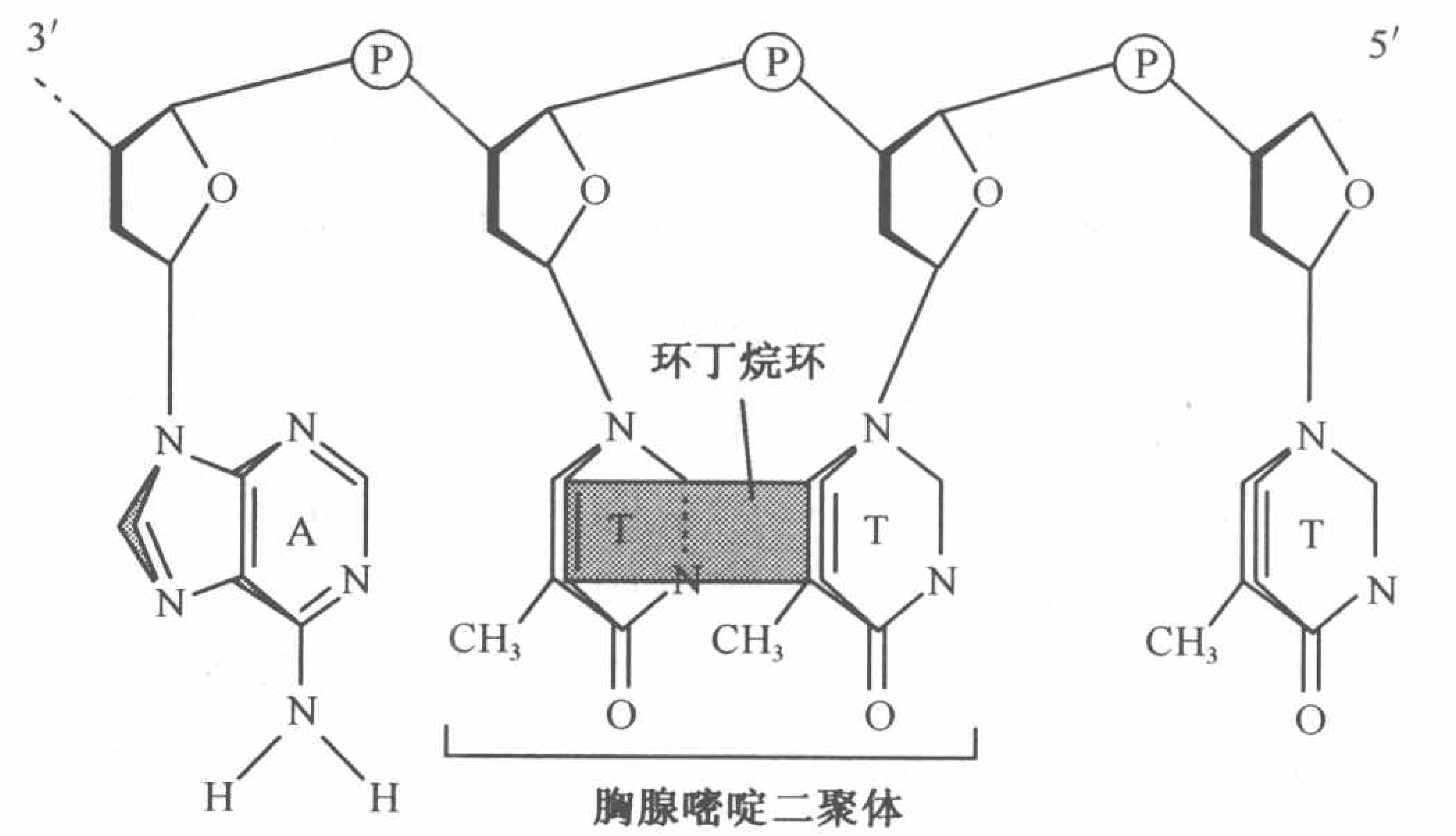


图 9-9 胸腺嘧啶二聚体。 紫外线引起毗邻胸腺嘧啶之间形成环丁烷环。

最后，γ 辐射和 X 射线（电离辐射）有很高的危险性，因为它们可使 DNA 双链断裂，这很难被修复。电离辐射能直接攻击（离子化）DNA 骨架中的脱氧核糖。另外，辐射能通过产生活性氧簇间接攻击（如上所述），因为后者能与脱氧核糖亚基反应。因为细胞需要完整的染色体来复制其 DNA，所以电离辐射用于在癌症治疗中治疗性地杀死快速增殖细胞。某些抗癌药物，如争光霉素（bleomycin）也能引起 DNA 的断裂。电离辐射和像争光霉素那样的试剂，因能导致 DNA 断裂，所以被认为具有断裂性（clastogenic）（源于希腊语 *klastos*，其含义为“断裂的”）。

突变还可由碱基类似物和嵌入剂引起

突变还可由可替换正常碱基的化合物（碱基类似物，base analog）或能插入碱基间的化合物（嵌入剂，intercalating agent）导致复制错误而引起（图 9-10）。碱基类似物在结构上与正常碱基类似，但是其不同之处使其对细胞又非常危险。因此，碱基类似物与正常碱基的相似程度足以使其被细胞摄取，转化成核苷三磷酸并在复制中被插入到 DNA 中。但是，因为这些碱基与正常碱基在结构上存在差异，使其配对不准确，导致复制过程中频繁发生错误。最有致变性的一个碱基类似物是 5-溴尿嘧啶（5-bromouracil），为胸腺嘧啶的类似物。溴取代原子的存在使碱基可以通过烯醇互变异构体与鸟嘌呤错配（图 9-10a）。如我们在第 6 章所见，酮互变异构体的倾向性远远大于烯醇互变异构体，但是胸腺嘧啶比 5-溴尿嘧啶更是如此。

如我们在第 6 章对溴乙锭所讨论的那样，嵌入剂是含有几个多元环的扁平分子，能与 DNA 中同样扁平的嘌呤或者嘧啶碱基结合，就如在双螺旋中的碱基相互结合或堆积那样。嵌入剂，如原黄素（proflavin）、吖啶（acridine）和溴乙锭（ethidium），能导致一个甚或几个碱基对的缺失或插入。当基因中发生这样的缺失或插入的时候，如第 15 章我们讨论遗传密码的时候将会看到的那样，会对其信使 RNA 的翻译产生深刻的影响，因为它们将编码序列移出了正确的读框。

嵌入剂是怎样引起短的插入和缺失的呢？插入情况的一个可能性是，这些致变剂通过插入模板链碱基之间，导致 DNA 聚合酶在嵌入分子对侧插入一个额外的核苷酸（一个这种结构的嵌入大约能使两个碱基对之间的通常间距加倍）。相反，缺失的情况是由于嵌入分子的存在引起模板扭曲，这可能使聚合酶跳过一个核苷酸。

DNA 损伤的修复

如我们所看到的那样，DNA 的损伤有两种后果。有些类型的损伤，如胸腺嘧啶二聚体或 DNA 骨架的切口和断裂，使得复制或者转录受阻。其他类型的损伤产生改变了的碱基，在复制上没有立即产生结构性后果，但是引起了错配，这可在复制之后导致 DNA 序列上永久性的改变。例如，通过脱氨基作用胞嘧啶转化成尿嘧啶产生 U：G 错配，在一轮复制之后，成为一条子染色体上的 C：G 到 T：A 的转换突变。这些考虑解释了为什么细胞进化出完美的机制以保证在损伤阻碍复制或者产生突变之前就识别并修

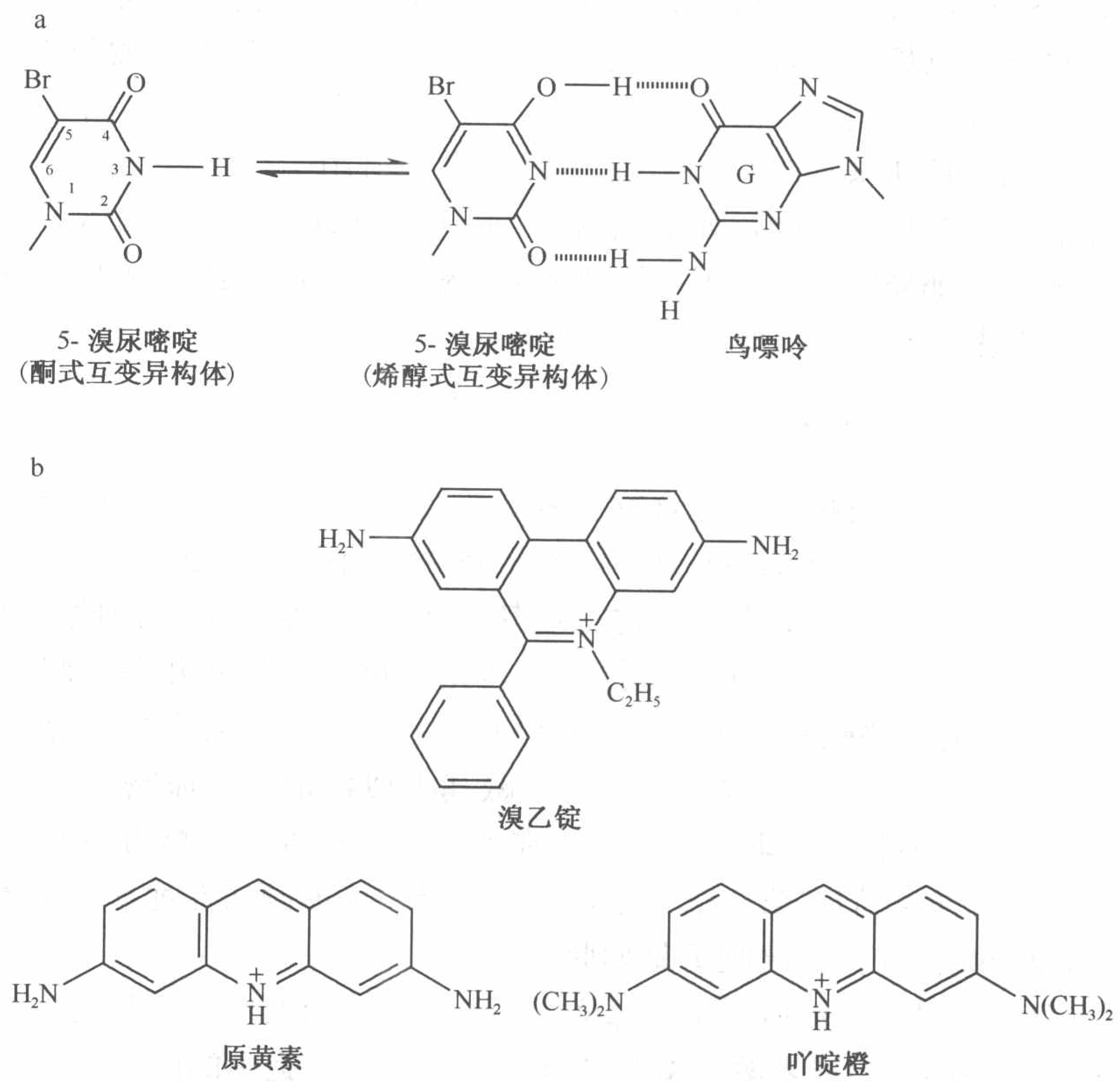


图 9-10 可引起 DNA 突变的碱基类似物和嵌入剂。 a. 胸腺嘧啶的碱基类似物，5-溴尿嘧啶能与鸟嘌呤错配。b. 嵌入剂。

复损伤。缺少这样的机制，细胞不能耐受多长时间。

在这一部分，我们探讨 DNA 损伤修复的系统（表 9-1）。这些系统最直接的方式是（指真正的修复）修复酶简单地将损伤逆转（撤销）。一个更精致的步骤涉及剪切修复系统（excision repair system），此处受损的核苷酸没有被修复，而是从 DNA 上被除去。在剪切修复系统中，另一条未受损的链作为模板，以便 DNA 聚合酶重新掺入正确的核苷酸。如我们将见到的，有两种剪切修复，一种仅仅将受损的核苷酸去除，另一种将包含损伤的一小段单链 DNA 去除。

表 9-1 DNA 修复系统

类型	损伤	酶
错配修复	复制错误	*E. coli* 中 MutS、MutL 和 MutH，人类中 MSH、MLH 和 PMS，
光激活作用	嘧啶二聚体	DNA 光解酶
碱基切除修复	受损的碱基	DNA 糖基化酶
核苷酸切除修复	嘧啶二聚体 碱基上的大加合物	*E. coli* 中 UvrA、UvrB、UvrC 和 UvrD，人类中 XPC、XPA、XPD、ERCCI-XPF 以及 XPG
双链断裂修复	双链断裂	*E. coli* 中 RecA 和 RecBCD
移损 DNA 合成	嘧啶二聚体或脱嘌呤位点	Y 家族 DNA 聚合酶，如 *E. coli* 中的 UmuC

但是，更精致的系统是重组修复（recombinational repair），当 DNA 两条链都受损（断裂）时采用这种修复。这种情况下，一条链不能作为另一条修复的模板。因此，在重组修复［也叫双链断裂修复（double-strand break repair）］中，序列信息从染色体第二条未受损的拷贝上找回。最后，当 DNA 聚合酶的复制进程被受损碱基阻碍的时候，一个特殊的移损（translesion）聚合酶以一种不依靠模板和新合成 DNA 链之间碱基配对的方式经过受损位点进行拷贝。因为移损合成不可避免地具有很高的错误易发性（致变性），所以细胞在迫不得已的情况下才采用此机制。

DNA 损伤的直接逆转

损伤简单逆转修复的一个例子是光激活作用（photoreactivation），光激活作用直接逆转紫外辐射造成的嘧啶二聚体结构。光激活反应中，DNA 光解酶从光线中捕获能量并将其用于断裂连接毗邻嘧啶的共价键（图 9-11）。换句话说，受损碱基被直接修复了。

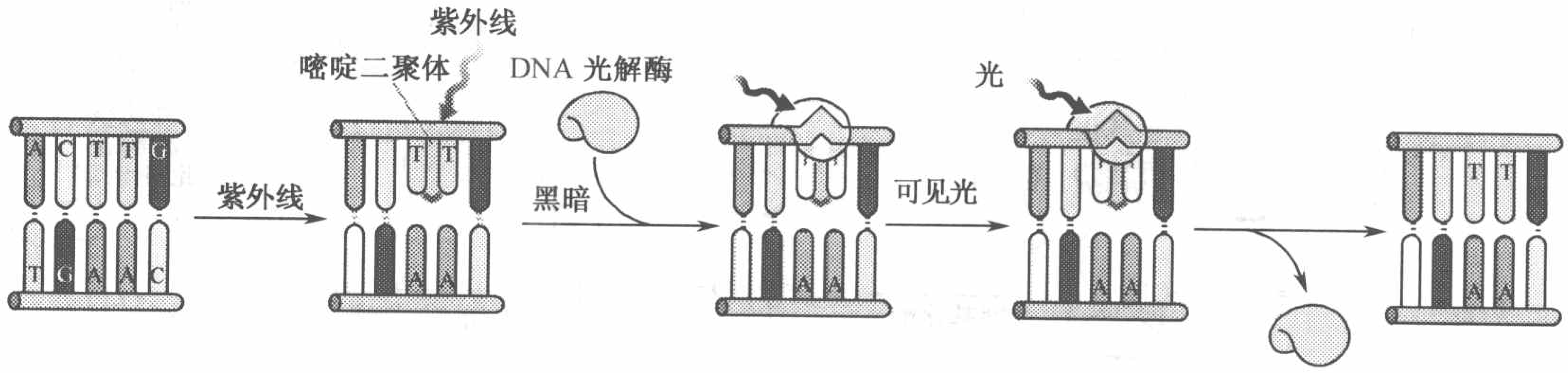


图 9-11 光激活作用。紫外线照射产生胸腺嘧啶二聚体。当被光照射时，DNA 光解酶将二聚体形成的环打断，恢复成两个胸腺嘧啶残基。

直接逆转的另一个例子是甲基化碱基 O^6-甲基鸟嘌呤上的甲基被去除（见上文）。在此例中，甲基转移酶把甲基基团转移到其自身的一个半胱氨酸残基上，从而使之从鸟嘌呤残基上除去（图 9-12）。对细胞来说这个代价很昂贵，因为甲基转移酶不具有催化性，一旦接受一个甲基基团就不可再次被使用。

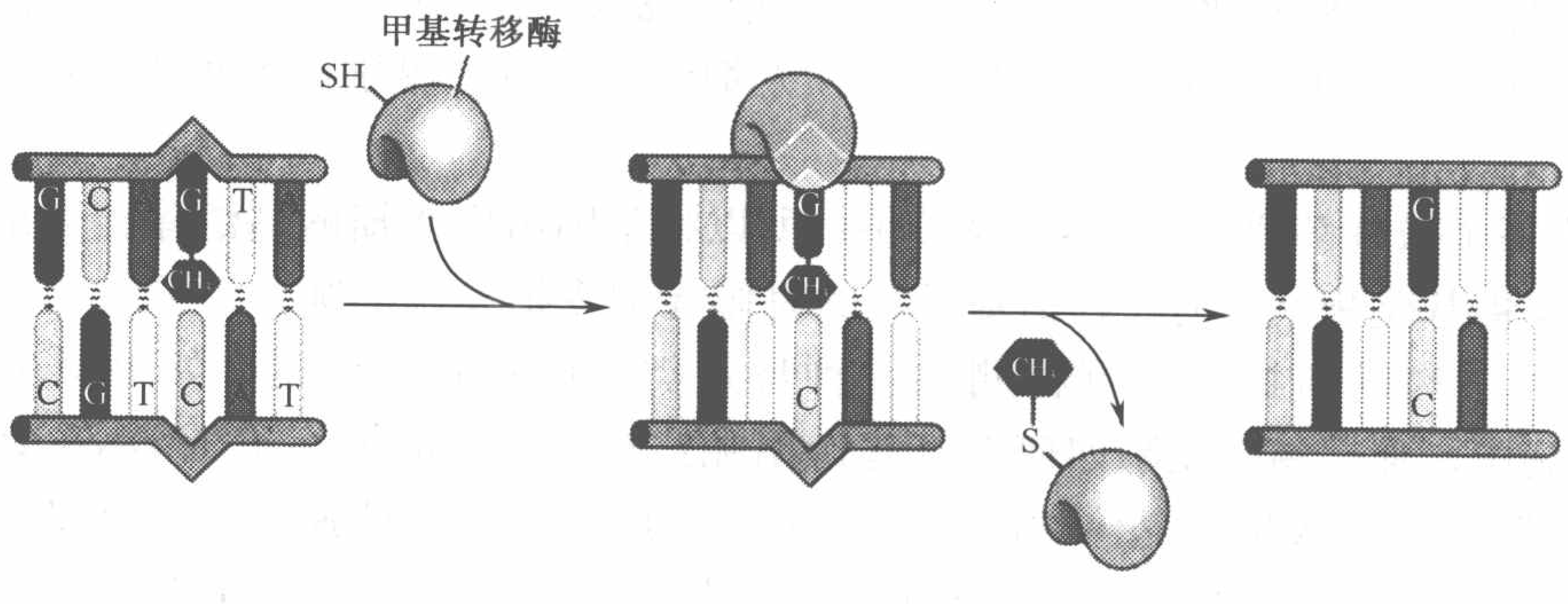


图 9-12 甲基基团的去除。甲基转移酶催化 O^6-甲基鸟嘌呤的甲基基团转移到该酶的一个半胱氨酸残基上，从而在 DNA 中恢复成正常的 G。

碱基切除修复酶通过碱基弹出机制除去受损碱基

DNA清除受损碱基最普遍的方法是通过修复系统将已变化的碱基除去并替换掉。两种主要的修复系统是碱基切除修复（base excision repair）和核苷酸切除修复（nucleotide excision repair）。在碱基切除修复中，一种称为糖基化酶（glycosylase）的酶识别并通过水解糖苷键除去受损碱基（图9-13）。在随后的核酸内切步骤中，产生的脱碱基戊糖从DNA骨架上去除。核酸内切性切割还除去自发水解产生的脱嘌呤和脱嘧啶的戊糖。受损核苷酸从骨架上完全除去以后，一种修复DNA聚合酶和DNA连接酶用未受损的链作为模板重建一条完整的链。

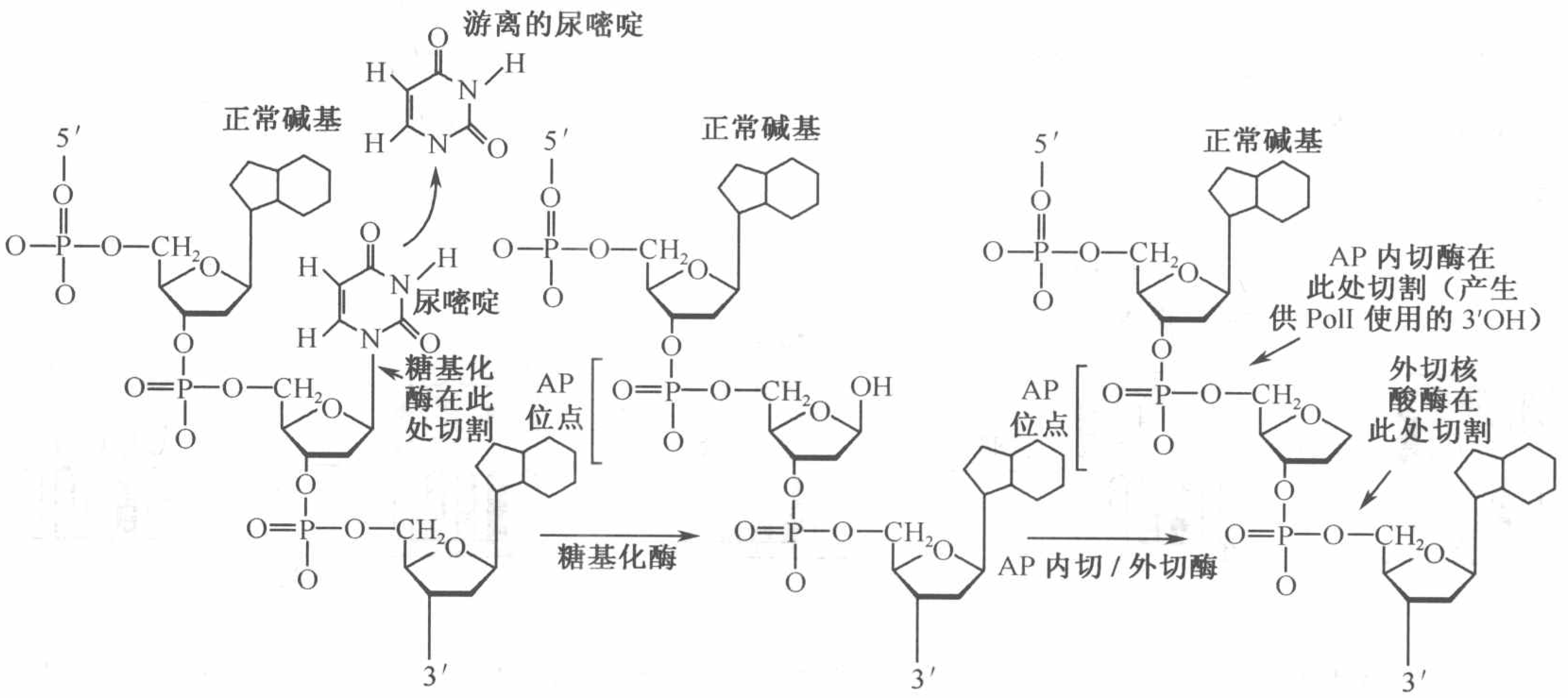


图9-13 碱基切除途径：尿嘧啶糖基化酶反应。 尿嘧啶糖基化酶水解糖苷键以使尿嘧啶从DNA骨架上脱离，并留下一个AP位点（脱嘌呤位点，或者在此情况下是脱嘧啶位点）。AP核酸内切酶在AP位点的5′位置上切断DNA骨架，留出3′-OH；核酸外切酶在AP位点的3′位置上切割，留出5′磷酸，产生的缺口被DNA聚合酶I填补。

DNA糖基化酶为损伤特异性的，细胞中有多种具有不同特异性的DNA糖基化酶。因此，一种特殊的糖基化酶识别尿嘧啶（胞嘧啶脱氨基作用产生），另一种负责除去氧代鸟嘌呤（oxoG，鸟嘌呤氧化的结果）。在人细胞核中总共鉴别出8种不同的DNA糖基化酶。

因为每个碱基都埋藏在DNA螺旋中，所以将基因组中受损碱基清除出去是一个难题。DNA糖基化酶是如何在扫描基因组的时候检测受损碱基的呢？证据表明，这些酶沿DNA的小沟纵向扩散直到检测出一种特定类型的损伤。但是，如果碱基包埋在DNA螺旋中，那么糖基化酶如何对其起作用呢？对此谜惑的解答源于DNA具有令人惊叹的柔韧性。对X射线晶体学的研究揭示了受损的碱基是被弹出的，所以它突出于双螺旋之外，此处它正好位于糖基化酶的特异性口袋中（图9-14）。有趣的是，双螺旋仅需中等的结构扭曲就能使碱基弹出，所以碱基弹出的能量消耗不会太大（第6章图6-8）。然而，糖基化酶沿着DNA扩散的时候，不可能将每一个碱基都弹出来进行检查

异常。因此，这些酶进行受损碱基扫描的机制仍然不为人知。



图 9-14 DNA 糖基化酶复合体的结构。酶显示为灰色，DNA 显示为紫色。图中受损的碱基——氧代鸟嘌呤显示为红色，从螺旋中弹出，进入酶的催化中心。（Bruner S. D.，et al. 2000. *Nature* 403：859-866.）镜像图用 BobScript、MolScript 和 Raster 3D 制作。

如果受损碱基在 DNA 复制之前没有通过碱基切除而被除去会怎样呢？这是不是意味着将不可避免地产生突变？在倾向于和 A 发生错配的 oxoG 例子中存在一种防故障机制（图 9-15）。一种专门的糖基化酶可识别模板链上 oxoG 配对位置错误掺入 A 产生的 oxoG：A碱基对。然而在这种情况下，糖基化酶会除去 A。因此，修复酶把与 oxoG 配对的 A 认作突变，并除去虽没有受损但是却不正确的碱基 A。

另一个防故障系统的例子是除去与 G 配对的 T 的糖基化酶。如我们所见，5-甲基胞嘧啶自发的脱氨基作用能形成 T∶G 错配，这在脊椎动物的 DNA 中常常发生。因为 T 和 G 都是正常的碱基，细胞如何识别哪一个才是不正确的碱基呢？糖基化酶系统认为，T∶G 错配中的 T 是 5-甲基胞嘧啶脱氨基作用产生的，并选择性地除去 T 使其被 C 取代。

核苷酸切除修复酶在损伤两侧切割受损的 DNA

与碱基切除修复不同，核苷酸切除修复酶不能区分各种不同的损伤，而是识别双螺旋形状上的扭曲，例如，胸腺嘧啶二聚体或碱基上存在大的化学加合物时产生的扭曲。这种扭曲引发一连串的事件，导致含有损伤的一小段单链片段（或小区）被去除。去除修复在 DNA 上产生的单链缺口由 DNA 聚合酶用未受损的链为模板进行填补，从而恢复原始的核苷酸序列。

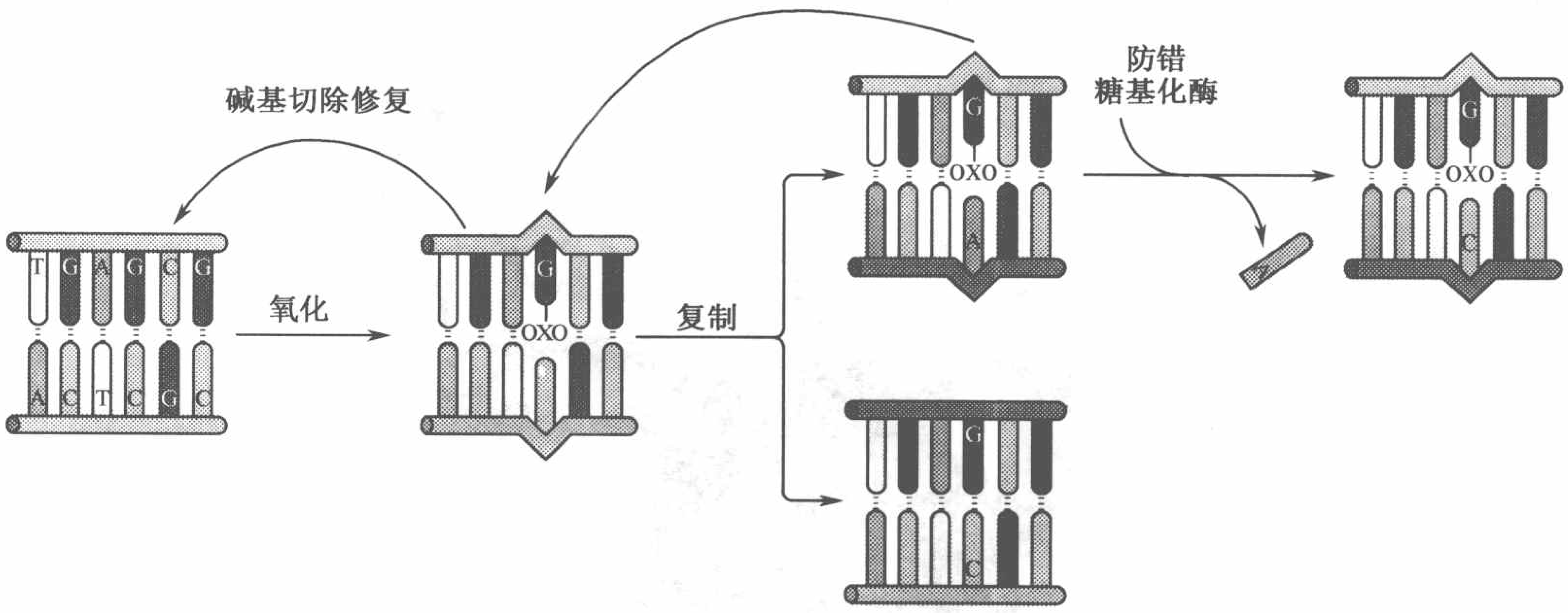


图 9-15 oxoG∶A 修复。鸟嘌呤氧化产生 oxoG。修饰后的碱基在 DNA 复制之前就能被 DNA 糖基化酶通过碱基切除途径进行修复。如果复制发生在 oxoG 被除去之前，致使 A 被错误插入，那么防错糖基化酶能除去 A，使 A 被 C 替换掉。这给 DNA 糖基化酶除去修饰的碱基提供了第二次机会。

E. coli 的核苷酸切除修复主要由 4 种蛋白质来完成：UvrA、UvrB、UvrC 和 UvrD（图 9-16）。UvrA 和 UvrB 组成的复合体扫描 DNA，其中 UvrA 负责检测螺旋中的扭曲，一旦遇到扭曲，UvrA 退出复合体。UvrB 使 DNA 变性，在损伤周围产生一个单链的凸起。然后，UvrB 募集 UvrC，UvrC 产生两个切口：一个位于损伤 5′侧的 8 个核苷酸处，另一个位于损伤 3′侧的 4 个或 5 个核苷酸处。切割产生长度为 12～13 个残基的单链 DNA 片段，此片段在 DNA 解旋酶 UvrD 的可作用范围之内。最后，DNA 聚合酶Ⅰ（Pol Ⅰ）和 DNA 连接酶填补产生的缺口。

较高等细胞中核苷酸切除修复的原理与 *E. coli* 中的基本相同，但是对损伤的检测、切除和修复系统更为复杂，涉及 25 个或者更多的多肽。其中 XPC 负责检测螺旋中的扭曲，这在 *E. coli* 中是 UvrA 的职能。与在 *E. coli* 中一样，DNA 被变性打开并在损伤周围形成一个凸起。凸起的形成涉及 XPA 和 XPD 蛋白的解旋酶活性（*E. coli* 中 UvrB 的对应物）以及单链结合蛋白 RPA。凸起在损伤部位的 5′侧产生的切割位点供核酸酶 ERCC1-XPF 切割，在 3′侧的供核酸酶 XPG 切割（代表 UvrC 的功能）。在高等细胞中，产生的单链 DNA 片段为 24～32 个核苷酸。与在细菌中一样，DNA 片段释放后产生的缺口被 DNA 聚合酶和连接酶填补。

如其名称所表明的那样，UVR 蛋白是紫外线损伤修复所需要的。*uvr* 基因突变株对紫外线敏感，缺少除去胸腺嘧啶-胸腺嘧啶二聚体和胸腺嘧啶-胞嘧啶加合物的能力。事实上，这些蛋白质广泛识别各种类型的巨大加合物并将之修复。核苷酸切除修复对人类也很重要。人类可患上一种名为着色性干皮病的遗传病，此病使患者对光线极度敏感并导致皮肤损伤，包括皮肤癌。已经鉴定出 7 种基因（称为 XP 基因），这些基因的突变将导致着色性干皮病。这些基因对应的蛋白质参与人类的核苷酸切除修复途径（如 XPA、XPC、XPD、XPF 和 XPG，见上文），这突出说明了核苷酸切除修复在紫外线损伤修复中的重要性。

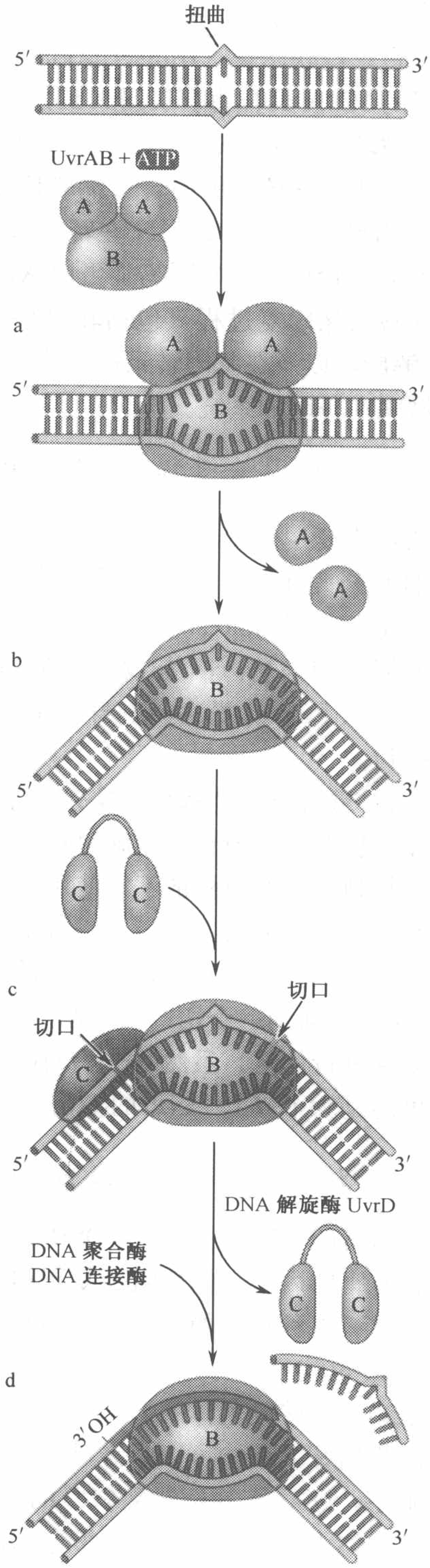


图 9-16 核苷酸切除修复途径。 a. UvrA 和 UvrB 扫描 DNA，寻找扭曲。b. UvrA 离开复合体，UvrB 在扭曲附近将 DNA 局部变性。c. UvrC 与 UvrB 形成复合体，在损伤的 5′侧和 3′侧产生切口。d. DNA 解旋酶 UvrD 将单链片断从双螺旋上释放出来，DNA 聚合酶 I 和连接酶对缺口进行修复和填补。（a～d经 Adapted from Zou Y. and Van Houten B. 1999. *EMBO J*. 18：4898，Fig 7. © Macmillan.）

核苷酸切除修复不仅能修复整个基因组中的损伤，而且当 RNA 聚合酶的转录过程被某个基因中发生的转录（模板）链的损伤所终止时，它还能拯救 RNA 聚合酶。此现象叫做转录耦联修复（transcription-coupled repair），包括将核苷酸切除修复蛋白募集到受阻的 RNA 聚合酶上（图 9-17）。转录耦联修复的意义在于它将修复酶集中于正在活跃地被转录的 DNA（基因）上。实际上，RNA 聚合酶在细胞中起到另一种损伤检测蛋白的作用。真核细胞中转录耦联修复的核心是通用转录因子 TFIIH。如我们将在第 12 章所看到的那样，TFIIH 在转录起始过程中解开 DNA 模板的螺旋。TFIIH 的亚基包括上面讨论的 DNA 解螺旋蛋白 XPA 和 XPD。所以，TFIIH 负责两个独立的职能：在核苷酸切除修复过程（包括转录耦联修复）中，有解链功能的解旋酶将损伤周围的 DNA 解旋，以及在基因转录过程中帮助打开 DNA 模板。原核生物中也存在将修复和转录耦联在一起的系统。

重组修复通过从未受损 DNA 中找回序列信息来修复 DNA 断裂

切除修复用未受损的 DNA 链作为模板来替代另一条链上已受损的 DNA 片段。细胞如何修复 DNA 中双螺旋两条链都断裂的双链断裂呢？这是由双链断裂修复（DSB）途径［double-strand break（DSB）repair pathway］进行的。此途径从姐妹染色体中取回序列信息。因为 DSB 修复途径在常规的同源重组以及 DSB 修复中具有核心作用，所以 DSB 修复途径本身就是一个重要的主题，我们将在第 10 章中对之进行详细地探讨。

DNA 重组还帮助修复 DNA 复制中的错误。设想复制叉在 DNA 中遇到没有被核苷酸切除修复纠正的损伤（如胸腺嘧啶二聚体），DNA 聚合酶有时将会放弃尝试越过此损伤进行复制。虽然模板链不能使用了，但是序列信息仍然能通过重组从复制叉的另一个子分子上取回（见第 10 章）。当重组修复完毕后，核苷酸切除系统又有一次修复胸腺嘧啶二聚体的机会。事实上，已经知道重组缺失的突变体对紫外线敏感。再考虑复制叉在 DNA 模板中遇到一个切口，复制叉越过切口将产生 DNA 断裂，这只能通过双链断裂修复途径进行修复。虽然一般我们认为重组是探讨序列新组合的进化策略，但它最初的功能可能是为了修复 DNA 的损伤。

DNA 中的 DSB 可以通过直接与断裂末端相连而修复

在所有的 DNA 损伤中，DSB 是对细胞最有害的。如果不修复，DNA 的断裂将引起多种有害的后果，例如阻断复制和引起染色体缺失，进而导致细胞死亡或肿瘤转化。细胞通常有多种重叠的途径来处理 DNA 损伤。那么细胞不仅仅依赖于 DSB 修复途径来处理 DSB 也就不足为奇了。正像我们所看到的以及在第 10 章中将要讨论的，DSB 修复途径依赖于姐妹染色体的 DNA 序列信息来修复受损的 DNA 分子。这是一种有效的策略，因为姐妹染色体可以为绕过断裂位点而精确地恢复原先的序列提供一个模板。在酵母细胞中，DSB 修复是断裂修复的一种主要的途径。但是，在细胞周期的早期，两个姐妹染色体由 DNA 复制产生之前又会发生什么呢？如果一个仍未复制的染色体发生了断裂，就没有姐妹染色体可以作为 DSB 修复途径的模板了。在这种情况下，另外一个

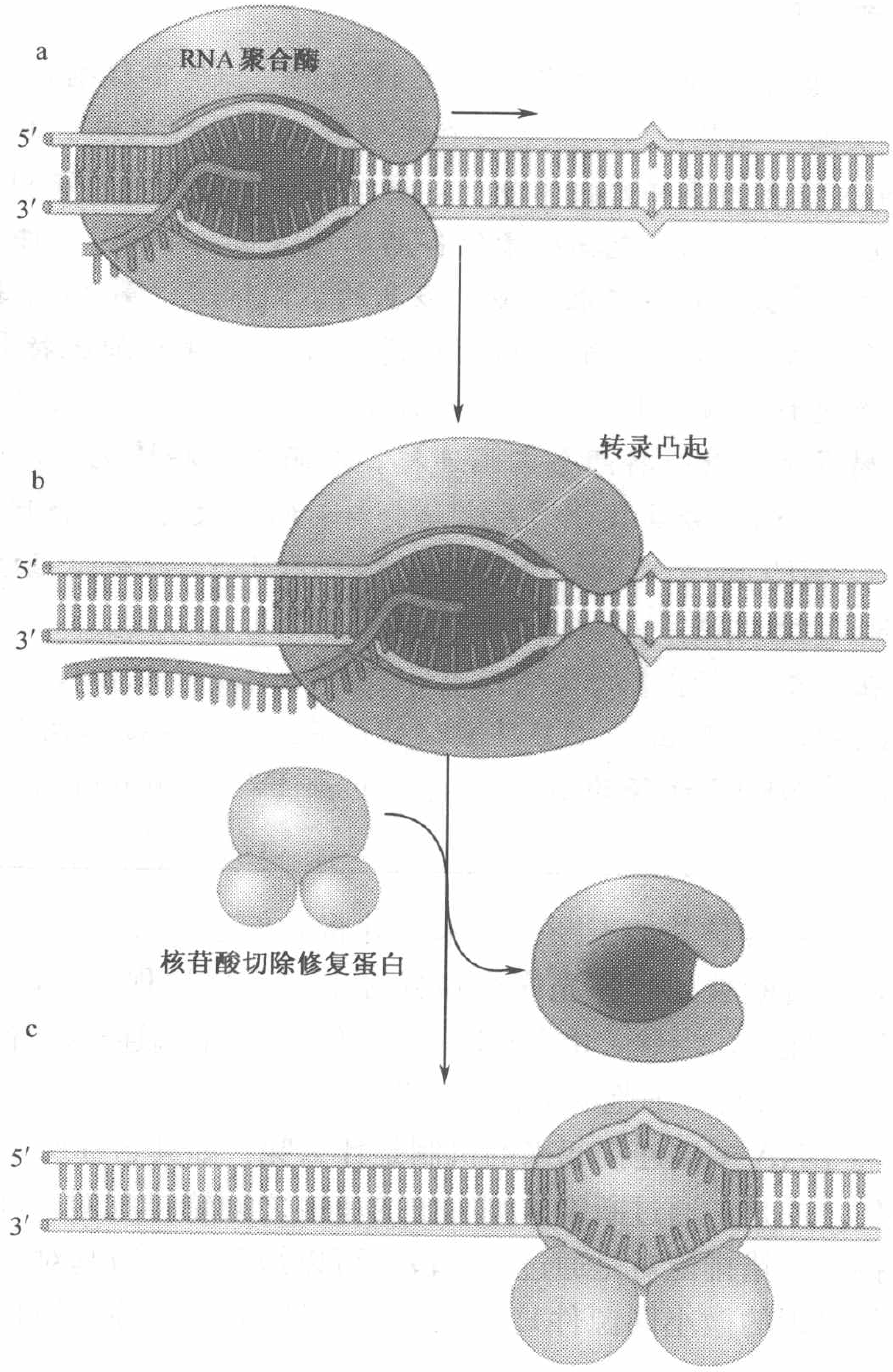


图 9-17 转录耦联的 DNA 修复。 a. RNA 聚合酶在损伤上游正常地转录 DNA。b. 当遇到 DNA 中的损伤时，RNA 聚合酶受阻，转录停止。c. RNA 聚合酶招募核苷酸切除修复蛋白到损伤部位上，然后从 DNA 上回退或者解离，以使修复蛋白能触及到损伤。（经 Zou Y. and Van Houten B. 1999. *EMBO J*. 18：4898，Fig 7. © Macmillan.）

称作非同源末端连接（Nonhomlogous end joining）（NHEJ）的系统就可以发挥作用了。NHEJ 是酵母中的备用系统，但是在高等细胞中它是修复断裂的一个主要的途径（框9-3）。

医学联系

框 9-3 非同源末端连接

NHEJ 能够修复暴露于外源有害物而产生的 DSB，如电离辐射以及细胞内在的损伤，如 DNA 复制失败。显然，NHEJ 也存在于所有正常的细胞内在的适应性免疫过程中。免疫系统产生大量多种多样的抗体分子，这些分子由所谓的轻链和重链组成。轻链和重链是在重组过程中产生的，包括大量储备的编码蛋白的 DNA 元件 V 和 J（在抗体的重链中是 D 片段）等大片段的连接，成为多肽的不同区域。第 11 章我们将这个过程称为 V（D）J 重组。V（D）J 重组的启动是通过在一个淋巴细胞特异的过程中引入 DNA 的断裂，这个过程还涉及 RAG1 和 RAG2 两个蛋白。一旦断裂产生，NHEJ 途径（并不是淋巴细胞特异的）就会将两个末端连接。然而在这种情况下，蛋白编码片段的末端并不是与它们原先的部分连接。这些末端会与新的片段连接起来形成重链和轻链中的合成编码序列。NHEJ 参与 V（D）J 重组的另一个例子，控制 T-细胞受体的免疫多肽的形成，将在第 11 章讨论。

强调 NHEJ 在人类生物学中的重要性是一些罕见的遗传综合征，它们的特征是对电离辐射和 DNA 损伤剂以及由 V（D）J 重组缺陷导致的免疫缺陷高度敏感。可见患有这些综合征的病人在 NHEJ 途径的 Artemis、Ligase IV 或 Cermummos-XLF 基因中发生了突变。

进行 NHEJ 的机器保护并加工断裂的末端并将它们连接在一起。由于断裂末端的序列信息丢失了，穿过断裂处的原先序列在异源末端连接过程中无法忠实地修复。因此，NHEJ 是有致变性的。当然，由 NHEJ 介导的 DNA 末端连接而导致的诱变要比将断裂的 DNA 留下而不修复对细胞的损伤小得多。

在 NHEJ 中将 DNA 末端连接起来的机制是什么呢？如其名称所示，NHEJ 不涉及同源重组。取而代之的是，通过断裂末端突出的单链之间错排配对，断裂 DNA 的两个末端直接相互连接。此错排配对是通过不片段（可以短至一个碱基对）互补碱基之间的匹配完成的（偶然发现的微小同源件）。单链尾巴用核酸酶去除，缺口用 DNA 聚合酶填补。

已经鉴定出了很多调控 NHEJ 的蛋白。至今，在哺乳动物细胞中已经发现了七个 NHEJ 途径的组分，它们的名字都比较难记：Ku70、Ku80、DNA-PKcs、Artemis、XRCC4、Cernunnos-XLF 和 DNA Ligase IV（图 9-18）。Ku70 和 Ku80 是 NHEJ 最基本的成分。它们能形成一个异源二聚体与 DNA 的断裂末端结合并招募蛋白激酶 DNA-PKcs。DNA-PKcs 接下来与 Artemis 形成复合体。Artemis 同时具有 5′到 3′外切核酸酶和潜在的内切核酸酶活性，可以被 DNA-PKcs 的磷酸化作用激活。这些核酸水解活性对断裂末端进行加工并为连接作准备。连接是由 Ligasa IV、XRCC4 和 Cernunnos-XLF 组成的复合物来执行的。

NHEJ 在真核生物中是普遍存在的，但是在细菌中很少发生。然而，有人在枯草芽孢杆菌（*Bacillus subtilis*）的孢子中发现了一个很有意思的特殊例子。枯草芽孢杆菌能

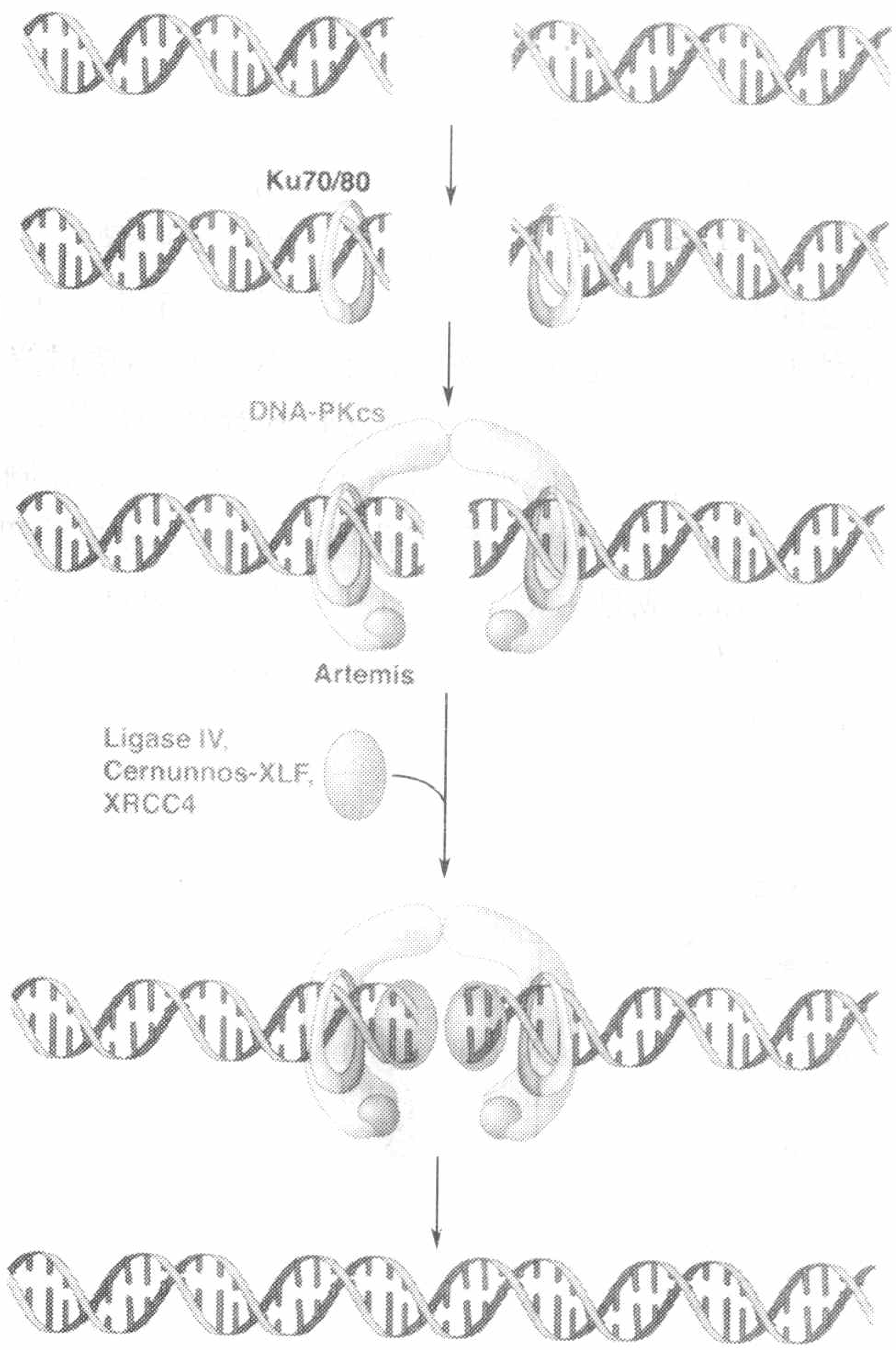


图 9-18 哺乳动物的 NHEJ 途径。Ku70 和 Ku80 组成的异源二聚体与断裂 DNA 的末端结合并招募蛋白激酶 DNA-PKcs。DNA-PKcs 紧接着招募拥有外切核酸酶和内切核酸酶活性的 Artemis 对断裂末端进行加工。最后，Ligase IV、XRCC4 和 Cernunnos-XLF 组成的复合物将两个裂末端相连。(Sekiguchi J. M. and Ferguson D. O. 2006. Cell 124：260-262 © Elsevier. 经许可重绘)

在孢子生成和成熟孢子组装蛋白的时候产生类 Ku 蛋白和一种 DNA 连接酶。Ku 蛋白和 DNA 连接酶代表了一个简单的二蛋白 NHEJ 系统，在孢子萌芽时修复 DNA 断裂。缺少这些蛋白的突变孢子对于干热（一种已知的引起 DNA 断裂的条件）非常敏感。在萌芽时期，加热突变的孢子并不能恢复生长，因为无法将热诱导的断裂重新连接。

萌芽时期的孢子主要依赖于 NHEJ 而不是 DSB 修复途径来修复断裂是有重要意义的。孢子只有一条染色体。因此，它们不能依赖于姐妹染色体作为修复突变的模板。有趣的是，孢子的染色体紧密卷曲形成一个典型的圆环圈结构可以保持断裂的 DNA 末端之间紧密地相邻。这种紧密地相邻可以使得正确的末端重连变得容易，甚至是染色体发生多处断裂的情况。枯草芽孢杆菌的孢子以及有关的细菌能在极端环境下比其他任何休眠的细胞更有效地生存。NHEJ 就是这种惊人技巧的基础。

移损 DNA 合成使复制越过 DNA 损伤继续进行

在迄今我们已经探讨的很多例子中，DNA 损伤是通过切除然后用未受损的模板重新合成来完成修复的。但是，这种修复系统不能十分有效地运行，有时正在复制的 DNA 聚合酶仍会遇到没有被修复的损伤，如嘧啶二聚体或脱嘌呤位点。因为这样的损伤是 DNA 聚合酶前进的障碍，所以复制机器必须试图越过损伤来进行复制，或被迫终止复制。即使细胞不能修复损伤，也有防故障机制使复制机器绕过受损的部位。这种机制就叫做移损合成（translesion synthesis）。如我们将看到的那样，这个机制有高度的易错性，易引入突变，但是移损合成使细胞避免了染色体不完全复制的更坏命运。

移损合成是被一类特化的 DNA 聚合酶催化的，此酶越过损伤部位直接合成 DNA（图 9-19）。在 *E. coli* 中移损合成是由 UmuC 和 UmuD′蛋白的复合体承担的。UmuC 是 DNA 聚合酶的一个独特家族的一员，它在很多生物中都有发现，称为 DNA 聚合酶的 Y 家族（图 9-20 和框 9-4）。

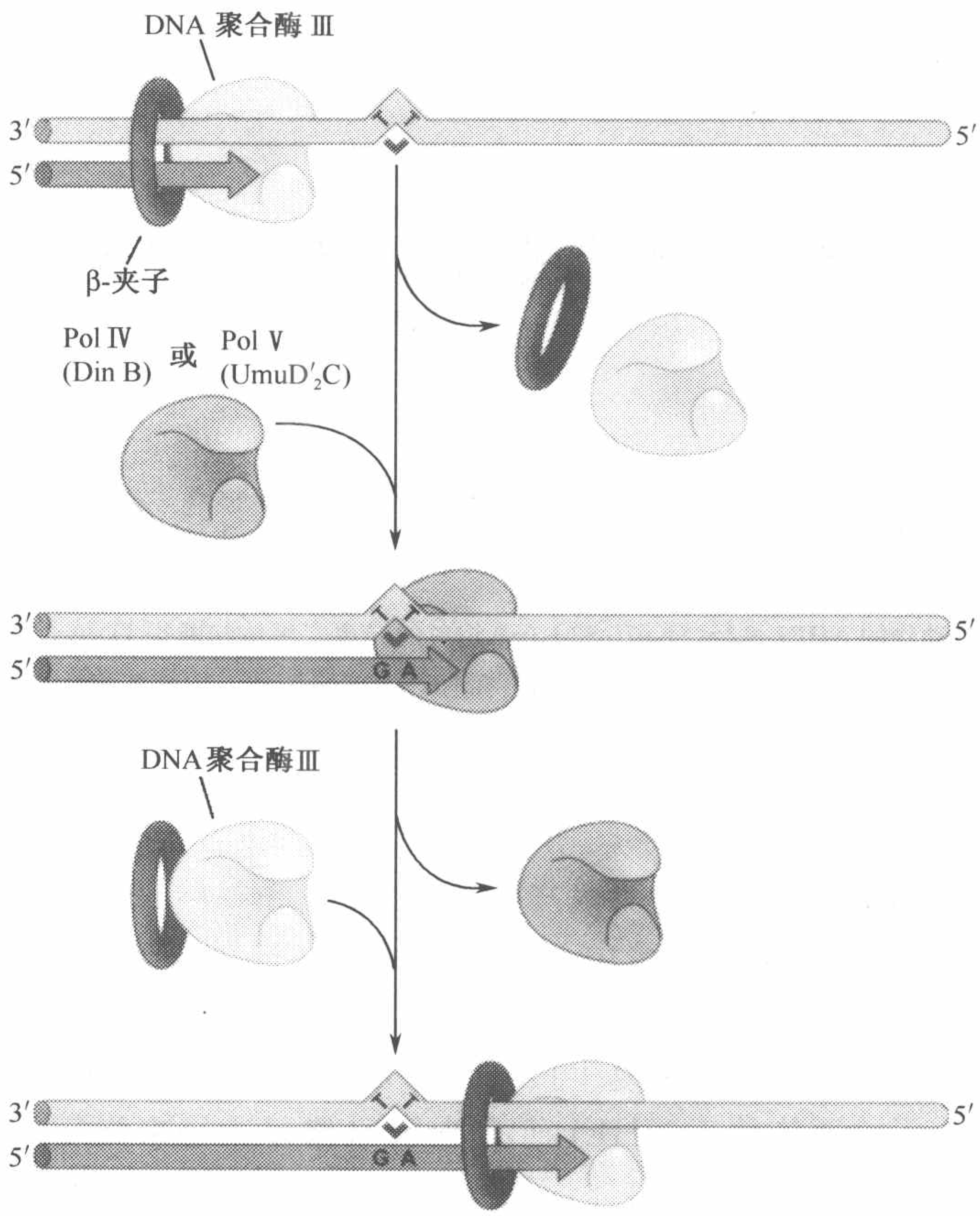


图 9-19　移损 DNA 合成。复制时一旦在模板中遇到损伤，DNA 聚合酶Ⅲ与其滑动夹一起从 DNA 上脱离下来并被移损 DNA 聚合酶取代。移损 DNA 聚合酶越过模板链（上者）上的胸腺嘧啶二聚体继续进行 DNA 的合成。然后移损聚合酶被 DNA 聚合酶Ⅲ替换。（Woodgate R. 1999. Genes Dev. 13：2191-2195，Fig. 1. © Cold Spring Harbor Laboratony Press. 经许可直绘）

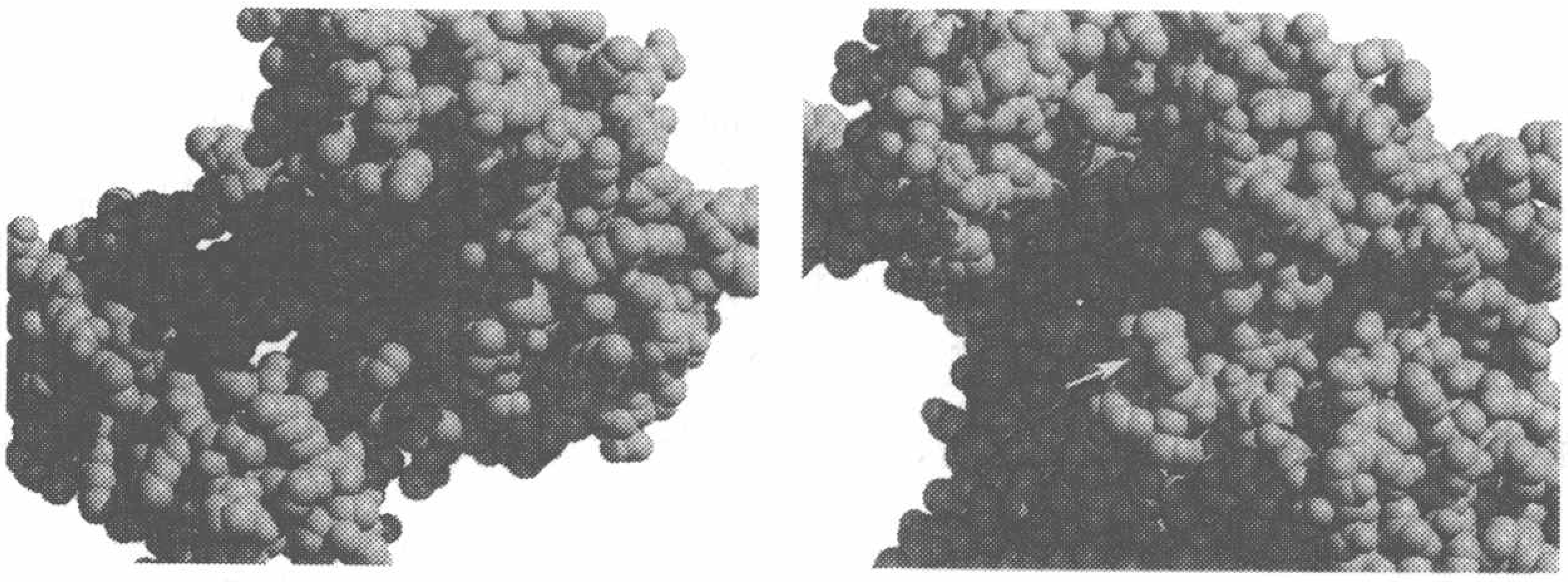
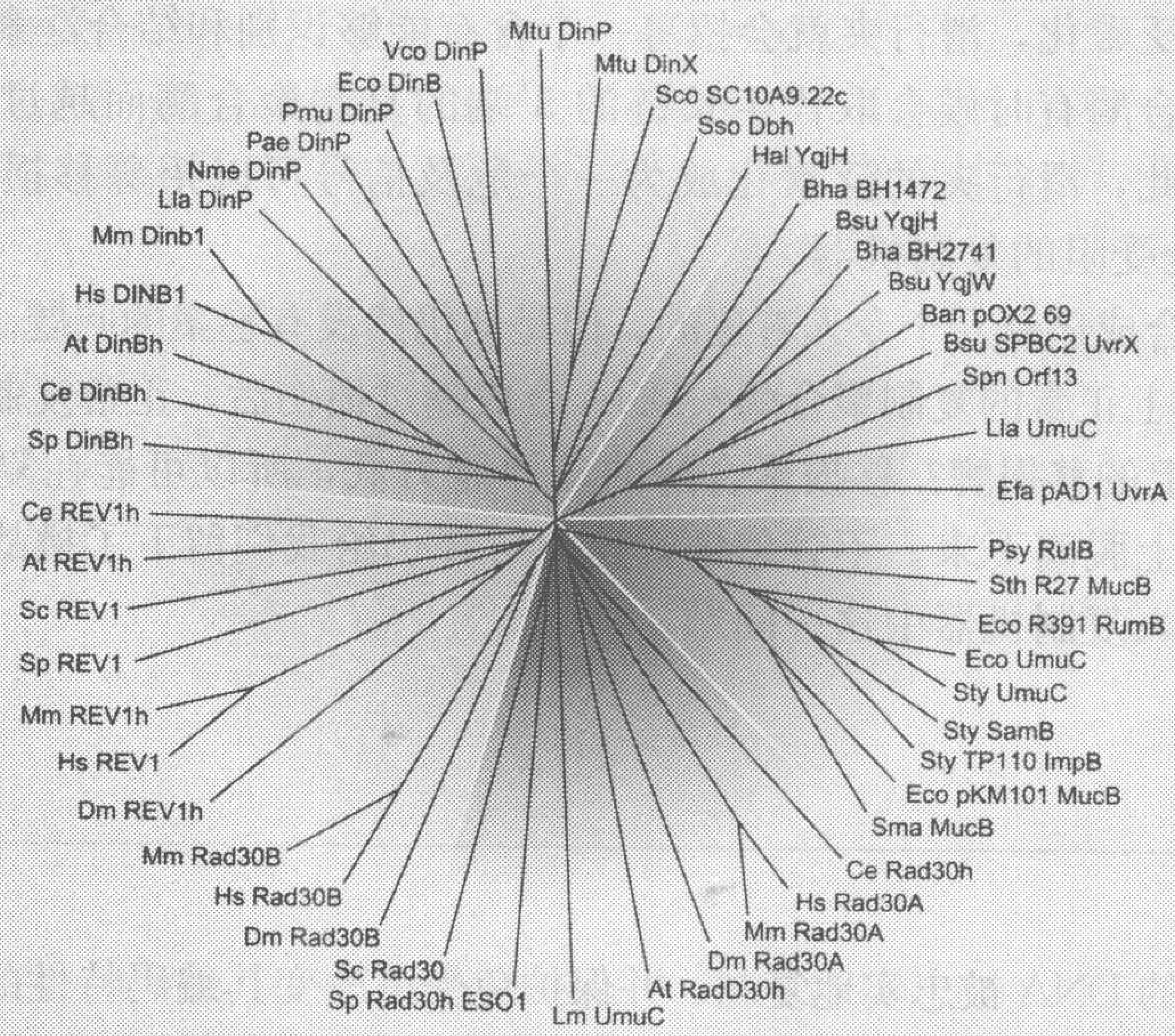


图 9-20 移损聚合酶的晶体结构。此处显示的是两种不同的 DNA 聚合酶的结构。左侧是个家族聚合酶（损伤旁路）的结构；右侧是噬菌体 T7 的典型的 DNA 聚合酶。注意 Y-聚合酶结构中活性位点更加开阔，而在右侧缺失的蛋白区域关闭了通道（黄色箭头指示）。掺进的核苷酸用红色表示，模板核苷酸用蓝色表示。（Y polymerase，Ling H. et al. 2001. Cell 107：91. PDB Code：1JX4. T7 polymerase，Doublié S. et al. 1998. Nature 391：251. PDB Code：1T7P.）镜像图用 MolScript，BobScript 和 Raster 3D 制作。

高级概念
框 9-4 Y 家族 DNA 聚合酶

DNA 聚合酶按照它们氨基酸序列的彼此相似性而划分成若干家族，在图中用不同的颜色表示（框 9-4，图 1）。最近证实 UmuC 和其他几种移损 DNA 聚合酶是一个庞大而独特的 DNA 聚合酶家族的创始成员，该家族被称为 Y 家族，在三种生命领域——细菌、古细菌和真核生物中都有发现。DNA 聚合酶的 Y 家族成员特征性地在未受损 DNA 模板上进行低保真度 DNA 合成，但是它能够越过阻碍 DNA 合成酶其他家族成员进行复制的 DNA 损伤。框 9-4 图 1 显示移损 DNA 聚合酶的 Y 家族系统进化树。



框 9-4 图 1 Y 家族 DNA 聚合酶的系统进化树。（Ohmori H. et al. 2001. Mol. Cell 8：7，Fig 1. © Elsevier）

这些聚合酶一个重要的性质就是，虽然它们是模板依赖性的，但是它们掺入核苷酸的方式不依赖于碱基配对。这解释了酶是如何越过模板链上的损伤合成 DNA 的。但是，因为酶没有从模板上“读取”序列信息。所以移损合成常有高易错性。试想在脱嘌呤或脱嘧啶例子中损伤位点并无碱基特异性的信息。移损聚合酶通过不按照碱基配对的方式插入核苷酸来越过损伤进行合成。虽然如此，插入的核苷酸仍不是随机的——有些移损聚合酶插入特定的核苷酸。例如，人的一个 Y 家族移损聚合酶成员在嘧啶二聚体对应位置上正确插入两个 A 残基。

因为它的高错误率，移损合成（类似于 NHEJ）被认为是最后可选择的求助系统。它能使细胞存活下来，否则可能对复制是灾难性的打击，但是付出的代价是高突变发生率。因此，在正常条件下，*E. coli* 中没有移损合成酶，此酶只有在对 DNA 损伤应答的时候才被诱导合成。因此，编码移损聚合酶的基因作为 **SOS 反应**（SOS response）途径的一部分来表达。损伤导致转录阻抑物（LexA 阻抑物）被蛋白质水解性分解，此转录阻抑物控制 SOS 反应涉及的基因的表达，包括 UmuC 和 UmuD（UmuD′无活性的前体）基因。有趣的是，同一途径还负责 UmuD 到 UmuD′的蛋白水解性转变。LexA 和 UmuD 的切割都是被称作 RecA 的蛋白质激发的，此蛋白质被 DNA 损伤产生的单链 DNA 激活。RecA 是双功能蛋白质，它还参与 DNA 的重组，将在第 10 章讨论。

接下来我们讨论一下移损聚合酶是如何进入 DNA 损伤位点处受阻的复制机器的。在哺乳动物细胞中，移损聚合酶是通过对滑行夹的化学修饰来进入移损修复途径的。在第 8 章中我们已经看到，滑行夹（也就是真核生物的 PCNA）将复制聚合酶锚定在 DNA 模板上。化学修饰是指将滑行夹与一个泛素化途径中的多肽（泛素）进行共价连接。在真核生物中，泛素化途径被广泛用于标记各种加工途径的蛋白，例如降解。泛素化途径能够引发移损合成最近发现越来越多，并且成为由泛素肽等标签蛋白控制的细胞过程之一。一旦泛素化，滑行夹就会招募一个含有能够识别和结合泛素的结构域的移损聚合酶。移损聚合酶有时候会取代生长链的 3′端的复制聚合酶而越过损伤位点进行生长链的延伸。因此，滑行夹的泛素化成为一个危及信号可以招募移损聚合酶来拯救在 DNA 损伤位点处受阻的复制机器。

最后，移损合成还有几个令人着迷的、但是却仍然没有答案的问题。移损酶如何取代 DNA 复制复合体上正常的复制聚合酶？DNA 合成越过损伤后，正常复制聚合酶怎样回来并取代在复制叉上的移损酶？移损聚合酶的延伸能力低，所以可能在复制通过损伤后不久它们就从模板上脱离下来。不管怎样，这个解释仍给我们留下了挑战，去解答正常加工酶如何重新进入复制机器中这一问题。

小结

只有当生物的 DNA 被忠实地复制，并保护免受可改变其编码特性的化学和物理损伤的时候，生物才能存活。复制精确度的限度和损伤修复的限度是由天然突变概率揭示的。因此，虽然单个碱基的错误概率可在 10 000 倍范围内变化，但是在正常情况下，一个普

通的核苷酸被复制 10^9 次，才有可能被错误改变约一次。复制精确性主要来自 DNA 聚合酶的复制模板的方式。正确碱基选择最初是根据互补配对进行的。DNA 聚合酶校正读码的活性增加了精确度。最后，在错配修复中，新合成的 DNA 链被一种酶扫描，然后该酶启动含不正确配对碱基 DNA 的替换过程。尽管有这些安全措施，但是所有类型的错误还是都发生了：碱基替换、小片段和大片段的插入和缺失、DNA 序列的大规模重排。

细胞含有大量用于修复 DNA 损伤的酶储备，这些损伤可能是致命的或改变 DNA 使其产生有破坏性的突变。有些酶直接逆转 DNA 损伤，如光解酶，其逆转了嘧啶二聚体的结构。更通用的策略是切除修复，即受损的片段被去除，并以未受损的链做模板合成新的 DNA 将之替换。在碱基切除修复中，DNA 糖基化酶和核酸内切酶只将受损的核苷酸除去；但是在核苷酸切除修复中，含有损伤的一小段单链 DNA 被除去。在 *E. coli* 中，切除修复是 UvrABC 核酸内切酶启动的，此酶在受损部位产生一个凸起，切去含有损伤的 12 核苷酸 DNA 片段。高等细胞用类似的方式进行核苷酸的切除修复，但是涉及的蛋白质数目更多，而且切除的单链 DNA 为 24～32 残基长。

最有害的损伤是 DNA 断裂。断裂是由重组或者 DSB 进行修复的，在这个过程中序列根据另外的一个同源的双螺旋进行复制而绕过断裂位点。如果在修复合成中没有模板可用，DNA 断裂会被 NHEJ 通过可能易错的方式将两端重新连接起来而修复。另一个易错的修复途径是移损合成。移损合成使得复制可以绕过阻断 DNA 聚合酶复制进程的损伤而继续进行。移损合成是受到一类特殊的和广泛存在的 DNA 聚合酶家族调控的。该家族能够以一种不准确且不依赖于碱基配对的方式进行 DNA 合成。

突变发生及其修复，因为永久地影响着生物继承的基因，并且因为癌症常常是体细胞突变引起的，所以令我们十分关注。

（张明　译　刘明旭　胡松年　刘韧　杨焕明　校）

参考文献

书

Friedberg E.C., Walker G.C., Siede W., Wood R.D., Schultz R.A., and Ellenberger T. 2005. *DNA repair and mutagenesis.* ASM Press, Washington, D.C.

Kornberg A. and Baker T.A. 1992. *DNA replication*, 2nd ed. W.H. Freeman, New York.

复制错误及其修复

Kunkel T.A. and Erie D.A. 2005. DNA mismatch repair. *Annu. Rev. Biochem.* **76:** 681–710.

DNA损伤的修复

Bridges B.A. 1999. DNA repair: Polymerases for passing lesions. *Curr. Biol.* **9:** R475–R477.

Citterio E., Vermeulen W., and Hoeijmakers J.H. 2000. Transcriptional healing. *Cell* **101:** 447–450.

Daley J.M., Palmbos P.L., Wu D., and Wilson T.E. 2005. Nonhomologous end joining in yeast. *Annu. Rev. Genet.* **39:** 431–451.

de Laat W.L., Jaspers N.G., and Hoeijmakers J.H. 1999. Molecular mechanism of excision nucleotide repair. *Genes Dev.* **13:** 768–785.

Drapkin R., Reardon J.T., Ansari A., Huang J.C., Zawel L., Ahn K., Sancar A., and Reinberg D. 1994. Dual role of TFIIH in DNA excision repair and in transcription by RNA polymerase II. *Nature* **368:** 769–772.

Kleczkowska H.E., Marra G., Lettieri T., and Jiricny J. 2001. hMSH3 and hMSH6 interact with PCNA and colocalize with it to replication foci. *Genes Dev.* **15:** 724–736.

Sekiguchi J.M. and Ferguson D.O. 2006. DNA double-strand break repair: A relentless hunt uncovers new prey. *Cell* **124:** 260–262.

第10章 分子水平上的同源重组

本章概要

•

DNA 断裂频发并启动同源重组

•

同源重组的模式

•

同源重组中的蛋白质机器

•

真核细胞的同源重组

•

交配型的转换

•

同源重组机制的遗传结果

所有的 DNA 都是重组 DNA。在同源染色体配对的减数分裂期第一次核分裂前，遗传信息的频繁交换最明显，使得染色体不断地被混合和重排。在配对阶段，染色体间发生遗传交换，这种现象通常被称为交换（crossing over），它是同源重组（homologous recombination）的结果之一，这种重组涉及染色体间 DNA 序列的物理交换。同一条染色体上的两个基因间发生交换的频率取决于它们之间的物理距离，相距越远越易发生交换。实际上，在以早期交换频率测量为基础构建的遗传图，通过揭示基因排列的固定及线性顺序规律，首次提供了染色体的结构信息。

但有时，基因的顺序会发生改变。例如，一种被称为转座子（transposon）的可以移动的 DNA 片段会随机地出现在染色体的不同位置上并促进 DNA 的重排，从而改变染色体的结构。导致“转座”及其他一些基因组重排的重组机制不同于同源重组，我们将在第 11 章中详细讨论这些现象。

同源重组是细胞的一个基本过程，由为专门合成专门调控的酶所催化。除了造成遗传多态性外，重组还能用细胞中未损伤的同源染色体 DNA 链，来修复因损伤而丢失的染色体序列。重组同时还提供一种机制，重启被停止或损伤的复制叉。此外，特殊类型的重组还能调控某些基因的表达。例如，通过交换染色体的特定部分，细胞可以把本来休眠的基因置于它们可以表达的区域。

除了解释遗传过程外，本章还要阐明重组的分子机制，该机制引领了基因操作方法的发展。这些基因操作方法包括常规应用的、在不同实验动物中产生“基因敲除”和“转基因”变异种的方法（见第 22 章）。依靠重组在整体生物的背景下去掉或引入基因，是确定基因功能非常有效的方法。

DNA 断裂频发并启动同源重组

DNA 经常发生双链断裂（Double-stnanded breaks，DSB）。如果断裂未得到修复，对细胞的后果是灾难性的。例如，*E. coli* 染色体上一个双链断裂如果无能力修复，对细胞是致命的。对大多数细胞而言，用于修复双链断裂的主要机制就是同源重组。有些细胞也可通过被称为非同源末端连接（NHEJ）的较简单机制进行修复。这个过程已在第 9 章中描述过。

细菌中，同源重组的主要生物学作用是修复双链断裂。这些断裂的 DNA 末端由多种方式形成（见第 9 章）。离子辐射和其他损害因子有时直接破坏 DNA 双链的骨架。许多类型的 DNA 损害可以经由干扰复制叉的进程而间接造成双链断裂。例如，一条 DNA 链上未修复的缺口能使经过的复制叉在此停止（图 10-1）。

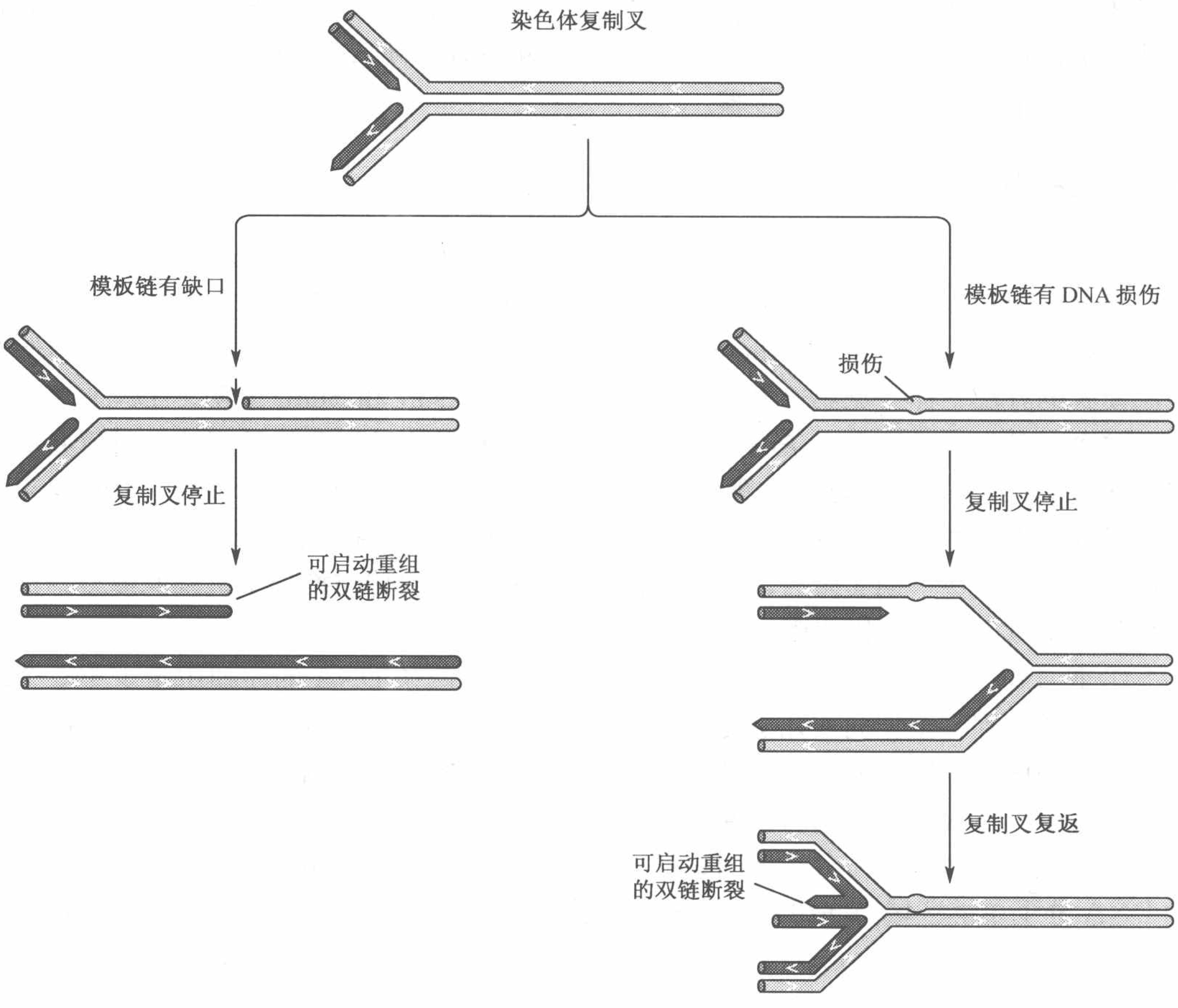


图 10-1 DNA 复制时模板上的损伤能导致双链断裂。模板链上单个缺口造成双链断裂易于理解（左图）。不过模板上的一个造成复制叉停止的损伤也能形成双链断裂（右图）。右图中，两条新合成的链（标为红色）可形成碱基互补链，导致复制叉向后退，这种结构有多种结局，断端可引发重组。

与此类似，一个使 DNA 不能作为模板的 DNA 链上的损伤可以导致复制叉停止。这种停止的复制叉的继续又有好几种途径（如复制叉后退、核酸酶分解等，图 10-1），并形成 DNA 双链断裂的末端。这些 DNA 断端会引发该 DNA 与其同源 DNA 形成重组，进而修复这个断裂。

同源重组除了能修复染色体 DNA 双链断裂外，还能促进细菌间的遗传物质交换。这种遗传物质交换发生在宿主细胞的染色体和经噬菌体转导或细胞接合而进入的 DNA 之间（见第 21 章）。在这些情况下，进入细胞的 DNA 为线性分子，提供能引发同源重组的 DNA 断端。

同源重组是真核细胞中修复 DNA 断裂和复制叉停止的关键步骤。染色体修复和复制重起是同源重组的主要功能，在高级生物的很多体细胞中是这样，在植物性生长的单细胞中也是这样。然后，在很多情况下，遗传交换和染色体稳定都需要同源重组。除此以外，同源重组还有其他作用，如下面将要提及的，重组是减数分裂时同源染色体配对的必要步骤。当细胞进入减数分裂时，产生一种特殊的蛋白质，介导 DNA 产生双链断裂从而引发重组（详情见下）。所以，虽然双链断裂有多种起因，DNA 上出现双链断裂仍是同源重组事件早期的关键步骤。

同源重组的模式

早期用原子同位素掺入 DNA 的巧妙实验，使人们对同源重组过程在分子水平上有了基本认识。这一实验与 Metthew Meselson 和 Frank W. Stahl 用以显示 DNA 半保留复制模式的方法一样（见第 2 章）。在实验中，Meselson 和 Stahl 证明了 DNA 复制产物中包含一条新合成链和一条旧链。与之相反，同样的实验过程揭示了这一所研究重组过程涉及 DNA 分子直接断裂和再结合。如下所述，现在我们知道 DNA 分子断裂和再结合是同源重组的主要特点。不过，重组也常常涉及 DNA 链局部的破坏（destruction）和再合成。在这些早期实验后，研究者提出了各种解释遗传交换分子机制的模型，这些模型都包括以下几个关键步骤：

① 两个同源 DNA 分子的联会。所谓同源指的是两条 DNA 的序列中至少 100 个碱基对以上的序列完全相同或几乎完全相同。除了这种高度的相似性，DNA 分子可有小部分区域的序列不同，如同一个基因的序列不尽相同的等位基因（allele）。

② 引入 DNA 断裂。这些断裂可进一步产生单链 DNA 区域。

③ 在两条重组 DNA 分子之间，形成碱基互补的短片段起始区。当来自于亲本分子的 DNA 分子单链区域与同源双链 DNA 分子互补链配对结合时，发生该配对过程，称为链侵入（strand invasion）。链侵入的结果是，产生了新的双链 DNA 区域；这种 DNA 经常带有一些错配的碱基称为异源双链螺旋 DNA（heteroduplex DNA）。

④ 链入侵之后两个 DNA 分子相互交叉的 DNA 链联系在一起。这个交叉的结构称为 Holliday 联结体（Holliday junction）。一个 Holliday 联结体可以通过配对碱基连续的解链和配对而沿着 DNA 移动。每次移动时，亲本 DNA 链上配对碱基断开，由相同的碱基配对形成重组中间体，该过程称为分支移位（branch migration）。

⑤ Holliday 联结体的剪切。在 Holliday 联结体处切断 DNA 重新生成两个分开的双螺旋 DNA，从而完成遗传物质的交换过程。这个过程叫做拆分（resolution）。在两个重组的 DNA 分子拆分时，哪两对 DNA 链在形成 Holliday 联结体时被切断，对于 DNA 交换程度有很大影响。

链入侵是同源重组的关键早期步骤

在损伤位点附近的链的加工处理过程详述如下。在同源重组中链，入侵是关键的一步，这种链入侵以及后续的在两个同源双链之间的互补链的配对在这些分子间会形成一个稳定的配对。这个过程也会引发两个亲本的 DNA 链的交换。下面我们将讲到，催化链入侵的酶称为链交换蛋白质（strand-exchange proteins），因为它们促进了这个关键反应。

将配对的两条同源双链染色体画成如图 10-2a 所示，对演示重源重组的各个步骤非常有用。尽管这些分子序列几乎完全相同，带有同一基因不同的等位基因（在图 10-2 中用符号 A/a、B/b、C/c 表示），这将帮助理解重组的结果。

重组是通过在每条 DNA 分子上同样区域引入一个双链断裂而始发的(图 10-2b)。断裂位置附近的 DNA 链可以从互补链的位置上释放出来，使之可供“侵入”并最终与同源 DNA 双链互补配对（图 10-2c)。在损伤位点附近的链加工处理以下会详述……

链侵入形成的 Holliday 联结体可通过分支移位沿着 DNA 移动，这种移位提高了交换 DNA 的长度。如果两个 DNA 分子并不完全相同，如同一基因不同的等位基因间常有的少量序列上的差异，发生在这些区域的分支移位形成的 DNA 双螺旋将带有 1 个或若干个序列错配（图 10-2d 中的 B 和 b 等位基因和旁边的放大图）。修复异源双螺旋 DNA 的错配有重要的遗传结果，本章的结尾处将提及这一点。

Holliday 联结体拆分是完成遗传物质交换的重要一步

完成重组要求 Holliday 联结体拆分，这个拆分过程是通过在交叉处切割 DNA 链；这个反应把两个重组的 DNA 分子分来，完成遗传物质的交换。图 10-3 显示了两个同源 DNA 双螺旋通过一个 Holliday 联结体连接在一起。拆分有两种不同的形式，可以形成两种不同的 DNA 产物，下面我们将做进一步的阐释。

图 10-3 显示了两个相似的双链 DNA 分子之间的交换产生的分支 DNA 上两对剪切点的位置。为了使这些剪切点易于观察，将 Holliday 联结体旋转，形成一个没有交叉的、十字路口形状的结构。序列相同且极性相同的两条 DNA 链将被切断。剪切位点的两种方式在图 10-3 中分别被标为 1 和 2。

在这个例子中，标为 1 的剪切位点发生在两条 DNA 链中，这两条链完全是由两个亲本 DNA 分子分别组成的（例如，蓝色实线和灰色的所示的链）。如果这些链现在被切割并共价的连接（第二个反应是由 DNA 连接酶催化的，后面会讨论），那么将产生的 DNA 分子的结构和序列如图的左下角。这些产物视为剪接（splice）重组产物，因为两个原先的双螺旋被剪接到了一起，来自亲本的 DNA 分子的区域与杂交双螺旋共价地连接到了一起。如下面所述的等位基因标签，产生的剪接产物导致了基因重排，这些

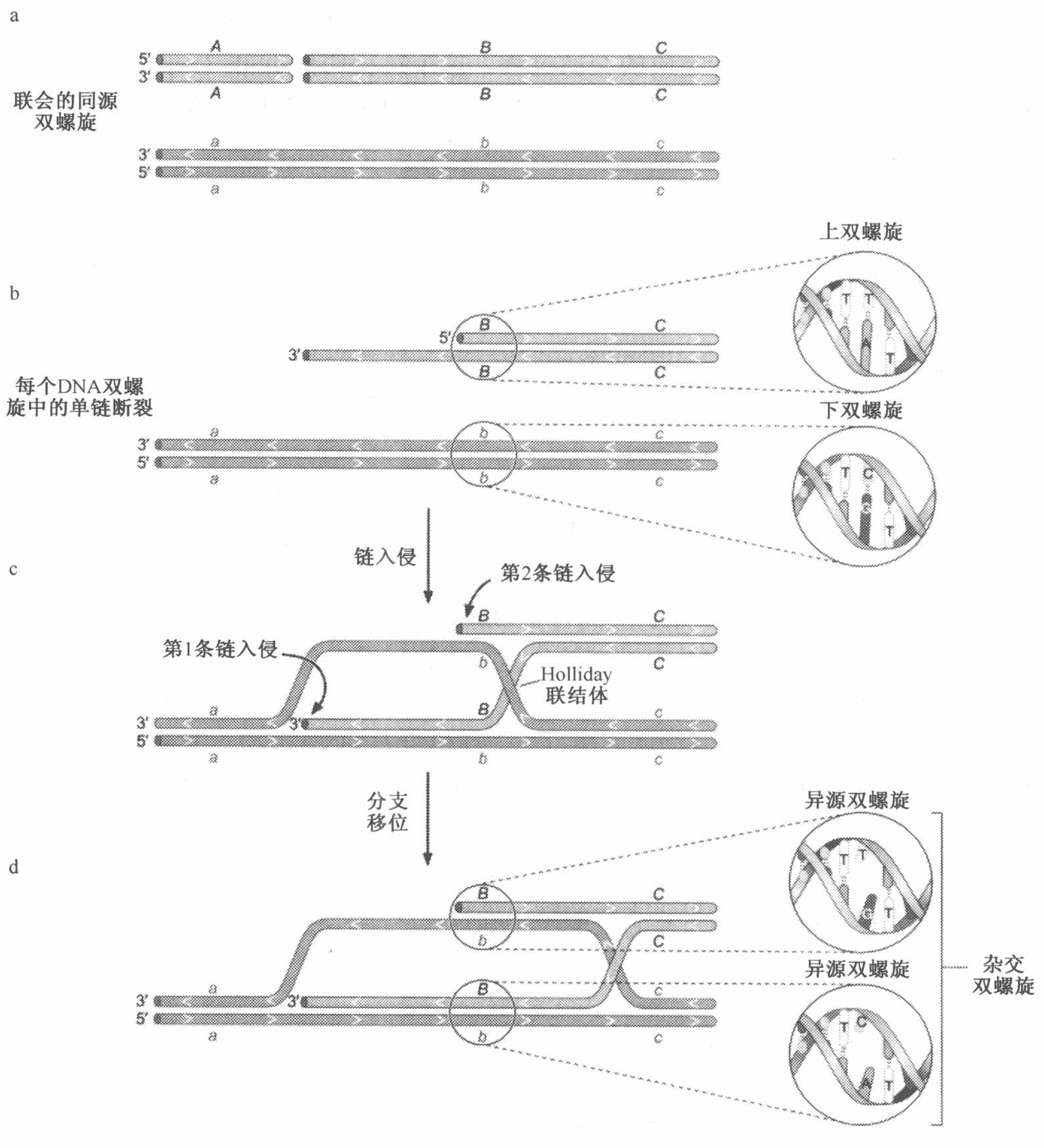


图 10-2 Holliday 模型中的分支移位。在 DNA 单链上的小箭头标示 5′→3′的方向。字母 *A* 和 *a*、*B* 和 *b*、*C* 和 *c* 分别代表了不同的等位基因。等位基因之间的序列差异不大。所以异源双链上同一基因座位，不同的等位基因之间就会形成错配。(如 d 图右边的放大图所示)

基因位于重组位点的两侧。因此这种类型的重组又被称为“交换产物” (crossover product)，在这种 DNA 分子中，交叉点位于 A 和 C 基因之间。

与之相反，Holliday 联结体中另一对剪切位点（图 10-3 中图标为 2）位于两条 DNA 链中并且含有两个亲本分子的序列区域（例如，蓝色和灰色片段）。在这些位点发生拆分和共价连接以后，产生了 DNA 分子包含一段杂合 DNA 链（补丁）。这些分子被称为补丁产物（patch product）。在这些产物中，重组并未产生两侧基因的重排（见图 10-3 中等位基因 A/a 和 C/c 等位基因标签）。这些分子又被称为非交换产物（noncrossover products）。影响拆分的位点和极性的因素将在下面讨论。

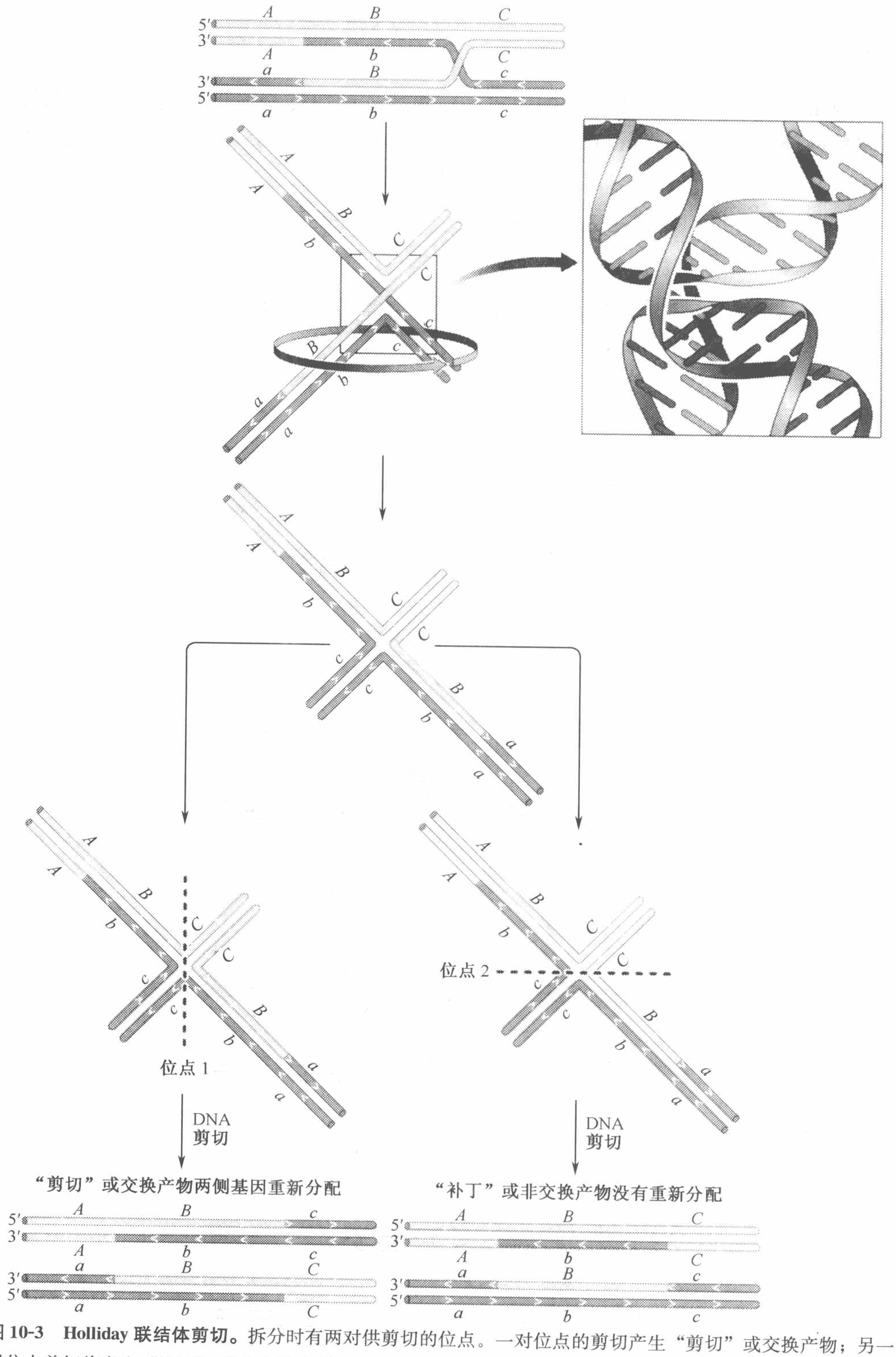


图 10-3 Holliday 联结体剪切。拆分时有两对供剪切的位点。一对位点的剪切产生“剪切”或交换产物；另一对位点剪切将产生“补丁”或非交换产物。小插图显示一个 Holliday 联结体的结构。请注意，小插图中的 DNA 碱基完全配对。

双链断裂修复模型描绘了许多重组事件

同源重组经常因 DNA 双链断裂而启动。描述这种遗传交换反应的常用模型是**双链断裂修复途径**（double-stranded break-repair pathway）（图 10-4）。在这个模型里，始

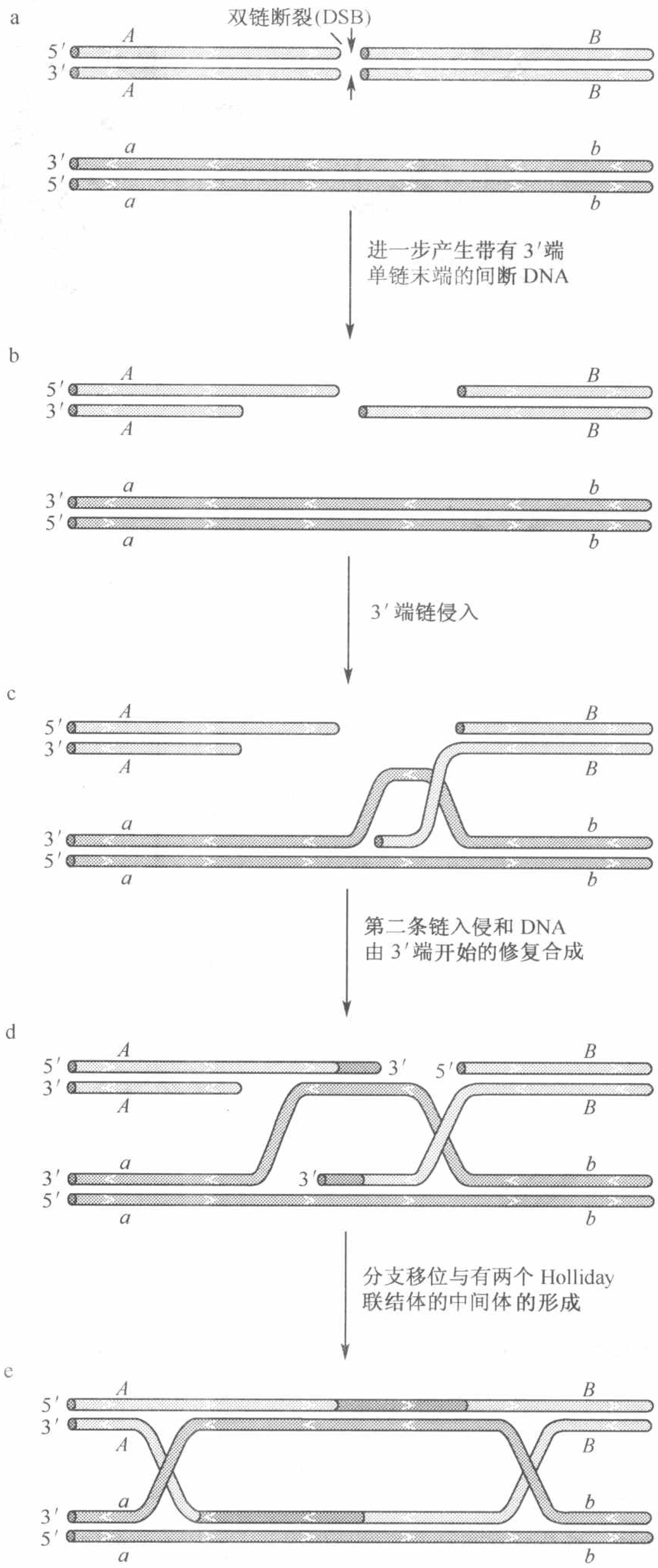


图 10-4 同源重组的双链断裂修复模型。图示产生有两个 Holliday 联结体的重组中间体的过程。

发事件是两条同源双螺旋 DNA 分子中的一条发生了双链断裂（见图 10-4a），而另外一条保持完整。两条 DNA 分子中的非对称始发断裂，导致双链 DNA 断裂模型的后期重组过程也是非对称的（即两条双螺旋的处理不同。在下文，我们将看到这一点）。

发生双链 DNA 断裂后，一种 DNA 裂解酶逐步降解断裂的 DNA 分子以产生单链 DNA 区段（图 10-4b）。该过程在断裂的 DNA 上产生了叫做单链 DNA 末端的以 3′端结尾的单链延伸体。有时双链断裂处的两条链都会被降解，有时只有 5′端的链被降解。

由上述过程产生的单链 DNA 末端这时侵入到未断裂的同源 DNA 双螺旋中（图 10-4c）。这张图示意了单链入侵，下一张图表示的是双链入侵。这两种情况中，每一条入侵链都是和另一条 DNA 分子中互补链配对。因为入侵链以 3′端结尾，它们可以作为合成新 DNA 的引物。以同源 DNA 双链中的互补链作为模板，从这些 3′断端开始延长，以重新合成断裂处被破坏的 DNA 片段（图 10-4d,e）。

如果原先的两条 DNA 分子序列在断裂附近并不完全一样（如上述的单碱基对替换），那么在重组的“双链断裂”修复中，序列信息就有可能丢失。如图 10-4 所示的重组事件，以标示的双链为模板合成 DNA 的结果导致 DNA（灰色）的序列信息丢失。在双链断裂修复模式下，这种单向的遗传物质交流有时会留下一个遗传标记——引发基因转换（gene conversion）事件，在本章结尾处我们将重述这一点。

双链断裂模式所产生的重组中间体——两个 Holliday 联结体，以分支移位的方式沿着 DNA 移动，并最终被拆分而结束重组事件。同样地，依据拆分 Holliday 联结体时所剪切的 DNA 链，可以决定最终 DNA 分子重组点两侧序列的基因是否发生重新配对（也就是导致交换）。如何拆分包含有两个 Holliday 联结体的重组中间体的不同方法，见框 10-1 的说明。

框 10-1　如何拆分有两个 Holliday 联结体的重组中间体

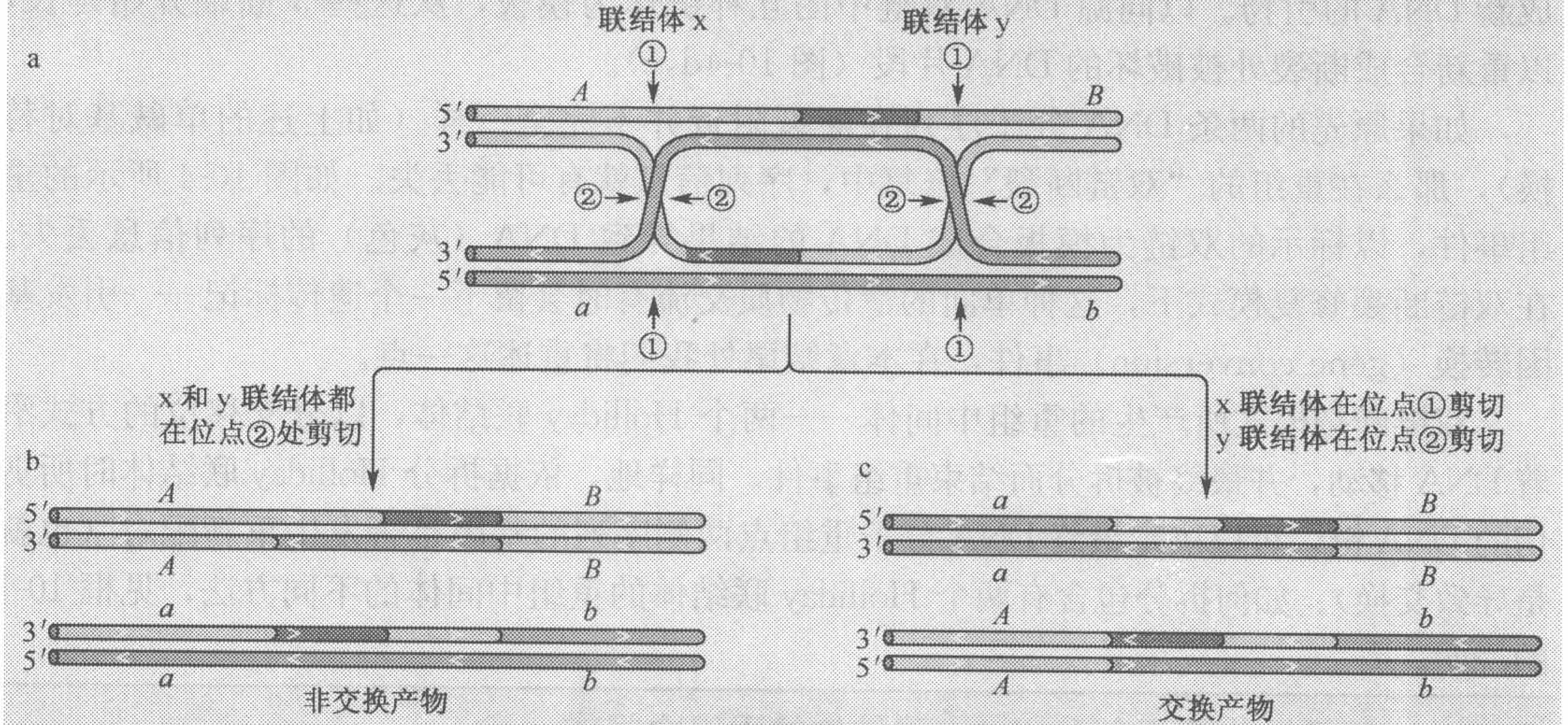
如何剪切重组中间体的 Holliday 联结体，对终产物 DNA 分子的结构有很重大的影响。终产物重组点两侧序列是重新分配（剪接/交换产物）还是不进行重新配对（补丁/非交换产物）决定于拆分如何进行。因为经双链断裂修复模型途径所产生的中间体有两个 Holliday 联结体，所以难于预见哪种产物是由哪种可能的 Holliday 联结体剪切组合而产生的。实际上，一种简单的方法可判定是否产生了交换或非交换产物。

下面以框 10-1 图 1 中分别标为 x 和 y 的两个联结体来解释拆分中间体的不同方式。对于每一个联结体，有两个可能的剪切位置（标为 1 和 2）。决定拆分是否产生交换或非交换产物的简单规则如下：如果两个联结体在同一个位置剪切，如都在 1 或都在 2，那么产生的是非交换产物。如图 1 中 b 所示，两个联结体都在 2 位置上剪切，产生了非交换产物。请注意等位基因 *A*/*B* 和 *a*/*b* 仍然在亲代染色体上同一 DNA 分子。两个联结体同时在 1 位上剪切也会产生非交换产物。

与之相反，如果剪切发生在两个联结体不同的位置，就会产生交换产物。如图 c 示，x 联结体在 1 位上剪切，y 联结体在 2 位上剪切。请注意现在基因 *A* 和 *b* 相连，而基因 *a* 则和 *B* 相连，两侧基因重新分配。如果 x 联结体在 2 剪切，y 联结体在 1 剪切，也会产生交换产物。

为什么这条简单的规则是真实的？为理解这一点，比较本图中的联结体和图 10-2 中的单 Holliday 联结体。单联结体在 1 位上剪切产生剪接产物，在 2 位上剪切则产生补丁产物。当你结合两个联结体的剪切结果时，便会有如下结论：

- 两个联结体都在 2 位上剪切，产生补丁产物（补丁＋补丁＝补丁，非交换产物）；
- 两个联结体都在 1 位上剪切，同样产生补丁产物（剪接＋剪接＝补丁，因为第二次剪接型拆分正好恢复了第一次剪切所造成的基因重排）；
- 若一个联结体在 1 位上剪切另一个在 2 位上剪切，就会形成交换产物（剪接＋补丁＝剪接），因为在 1 位上剪切形成的基因重排在对最终产物中得以保留。



框 10-1 图 1 双链分裂途径中间物的二种不同的拆分方式，亲代 DNA 分子连接，如图 10-4 所示。红色标示的 DNA 区域在重组中重新合成。

同源重组中的蛋白质机器

所有的生物都编码能催化 DNA 重组各个生化步骤的蛋白质。有些重组步骤中，同源蛋白质家族的成员在所有的生物中提供同样的功能。反之，另外一些重组步骤中不同生物由不同种类的蛋白质催化，但都产生同源的整体的结果。我们对重组机制最详实的理解来自于对大肠杆菌及其噬菌体的研究。所以我们首先着重描述主要通过双链断裂修复途径［又称 RecBCD 途径（RecBCD pathway）］促进大肠杆菌重组的蛋白质。真核细胞同源重组及其涉及的蛋白质将在后面的部分介绍。

表 10-1 列出了在细菌和真核细胞（如芽殖酵母是最深入理解的真核细胞）中催化关键重组步骤的蛋白质。这些蛋白质活性是完成双链断裂修复路径关键步骤所必需的。除了这些负责重组的蛋白质外，DNA 聚合酶、单链 DNA 结合蛋白、拓扑异构酶、连接酶等也在遗传物质交换过程中起着关键作用。

表 10-1　原核和真核细胞中催化重组步骤的因子

重组步骤	大肠杆菌中的催化蛋白	真核细胞中的催化蛋白
同源 DNA 联会及链侵入	RecA 蛋白	Rad51，Dcm1（减数分裂期）
引入双链断裂	无	Spo11（减数分裂期），HO（交配模式转换）
作用于 DNA 断端以产生用于侵入的单链	RecBCD（解旋酶/核酸酶）	MRX 蛋白（又称为 Rad50/58/60 核酸酶）
链交换蛋白组装	RecBCD 和 RecFOR	Rad52 和 Rad59
Holliday 联结体识别和分支移位	RuvAB 复合体	未知
拆分 Holliday 联结体	RuvC	可能是 Rad51c-XRCC3 复合体或其他蛋白质

请注意，尽管在 RecBCD 路径的重组中两个重组 DNA 分子中至少要有一个双链断裂，但在表 10-1 中没有列出大肠杆菌的引入双链断裂的蛋白质。如前所述，迄今为止尚未发现细菌中有引入双链断裂功能的特殊蛋白质。实际上，DNA 损伤和复制叉停止等事件是产生双链断裂（在染色体 DNA 上引发同源重组）的主要来源。另外在遗传物质交换的反应中，如噬菌介导的转导（我们将在第 22 章阐述），感染的 DNA 片断带有断裂的 DNA 末端。

下面将描述大肠杆菌的重组蛋白质，及其如何在重组过程中通过双链断裂修复路径行使功能，我们将根据这些蛋白质在反应过程中出现的顺序加以讨论。首先讨论的是 RecBCD 酶如何作用于 DNA 的双链断裂点进而产生单链区；然后，阐述产生链交换作用的 RecA 蛋白的结构和作用机制，当 RecA 蛋白结合在 DNA 单链上后，能发现 DNA 上的序列同源区并在这些同源区域间形成新的碱基对；接下来，讨论促进分支移位的 RuvA 和 RuvB 蛋白；最后，介绍拆分 Holliday 联结体的酶 RuvC。

RecBCD 解旋酶/核酸酶加工用于重组的 DNA 分子断裂

具有单链延伸端或“尾巴”的 DNA 分子是引发同源序列间链交换的合适底物。RecBCD 酶作用于断裂的 DNA 分子以产生单链 DNA 区域，RecBCD 也可协助链交换蛋白质 RecA 结合到单链 DNA 末端。另外，RecBCD 所拥有的多种酶活性提供给细胞一种“选择”，或是与进入细胞的 DNA 分子发生重组，或是降解它。

RecBCD 是由 3 个亚基所组成（*recB* 基因、*recC* 和 *recD* 基因的产物），并有 DNA 解旋酶和核酸酶活性。RecBCD 依靠 ATP 水解提供的能量，结合在 DNA 双链断裂处并沿 DNA 移动。在该蛋白质的作用下，DNA 被解旋，同时伴有或不伴有破坏一条或两条 DNA 链。RecBCD 蛋白的活性受控于一段特殊的 chi 位点的（cross-over hotspot instigator）DNA 序列。chi 位点由于能够提高同源重组的频率而被发现。

图 10-5 显示 RecBCD 作用于具有一个 chi 位点的 DNA 分子，使之受激活而进行重组的过程。RecBCD 蛋白通过双链断裂处进入 DNA，并在沿着 DNA 链移动的过程中解旋。RecB 和 RecD 亚基都是 DNA 解旋酶，利用 ATP 水解的能量解链 DNA 碱基对

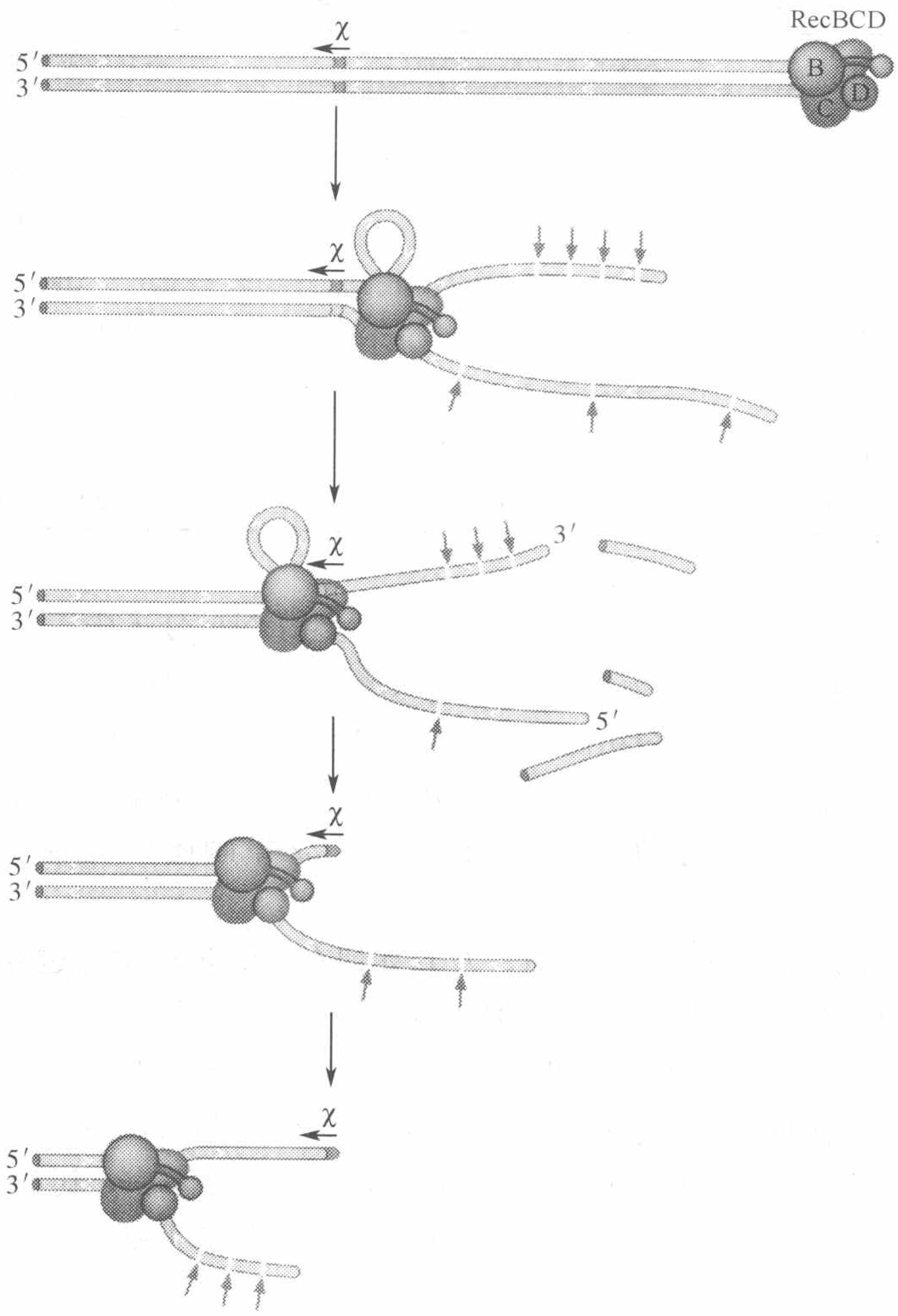


图 10-5　RecBCD 作用于 DNA 的步骤。请注意，RecBCD 蛋白能通过双链断裂中的任一个或两个断裂端进入 DNA 分子。但是 chi 位点的功能只有一个方向。如图示 DNA 分子中，chi 位点的方向使其只作用于从右向左移动的 RecBCD 酶。RecBCD 具有两个解旋酶：RecD 酶可沿着 DNA 5′端（图中下链）快速移动；RecD 酶则沿着 DNA 3′端（图中上链）慢速移动。因为这两个亚基以不同的速度移动，所以在解旋的过程中，会在上链堆积出一 DNA 单链环。当 RecD 到达 chi 位点时，脱落或失活。

（见第 8 章）。RecB 是个 3′-5′解螺旋酶并且有一个多功能的 DNA 核酸酶的结构域（见下）而 RecD 是-T5′-3′的解旋酶，RecE 具有识别 chi 位点的功能，RecBCD 的核酸酶活性经常在解旋的过程中剪切两条链从而降解结合的 DNA。

遇到 chi 位点后，RecBCD 的核酸酶活性发生变化。随着 RecBCD 移入 chi 位点之外的远端序列（相对于 RecBCD 酶进入的双链断裂），它不再依 3′→5′方向剪切 DNA，而是以更高的频率剪切另一条 5′→3′方向的 DNA 链。这种变化的结果使一条双螺旋 DNA3′单链延长到 3′端的 chi 位点。这种结构适于 RecA 蛋白的组装，并引发链交换（详情下述）。RecBCD 遇到 chi 位点后酶活性变化的分子机制，与 RecD 亚基的活性丧

失以及 DNA 通过 RecBCD 多亚基的变化方式有关。

与 DNA 结合的 RecBCD 复合体结构使我们更加深入地了解这三个亚基的功能和当遇到 chi 位点后活性如何变化的（见结构指南 10-1）。如图 10-6 所示，这个蛋白质复合体是一个三角形，双螺旋 DNA 是从三角体的上部进入蛋白质的。在这里，DNA 遇到一个称为“栓”的结构，这个结构是由 RecC 亚基突出来的，它的作用是切割双螺旋并且引领各自的 DNA 链到两个酶的驱动区。RecC 亚基引领 3′链进入 RecB 驱动区并使 5′链进入 RecD 驱动区。在这个情况下，RecC 虽然本身不是一个解旋酶，但提高了整个复合体的整体解旋活性的效率。酶中孔道的形成使得 3′DNA 尾巴恰好暴露于 RecB 亚基核酸酶活性位点处。因此在酶遇到 chi 位点前这条链就被高效地降解了。DNA 的 5′尾部离开 RecD 驱动区后同样会暴露于 RecB 核酸酶的活性位点处，但是它的降解效率低于 3′端尾部，因为它必须与更易被降解的 3′尾部竞争。然而，一旦遇到一个 chi 位点，这种情况就会发生了变化。RecC 识别并紧紧结合此 DNA 位点，并且一旦 chi 位点的 3′端被结合后，3′尾部就不能进入核酸酶，也就阻止了 3′端的降解。这样通过活性位点处竞争者的移除促进了 5′尾部的降解。

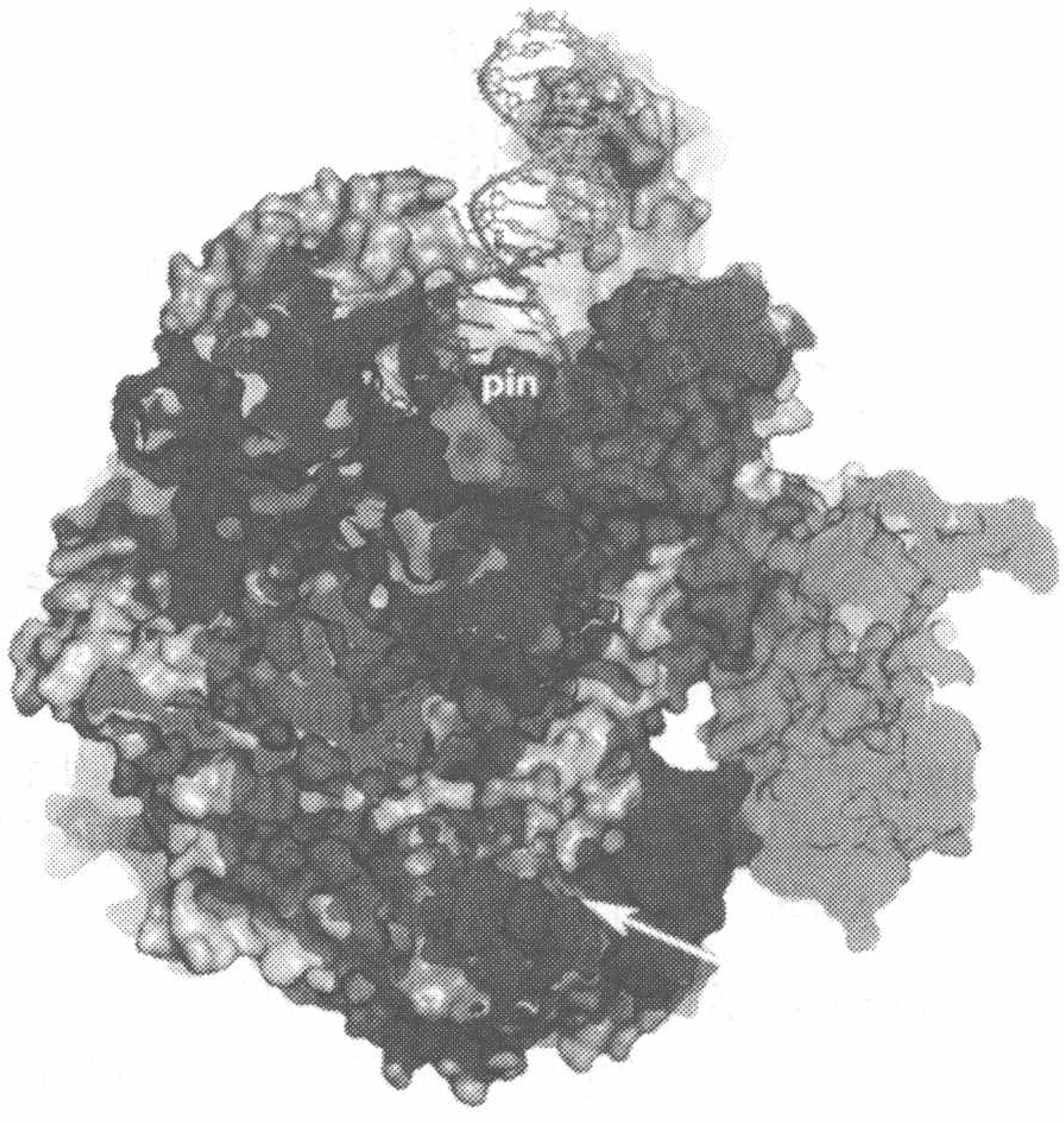


图 10-6　RecBCD-DNA 复合物的结构。这 RecB 标为蓝色，RecC 为红紫色，RecD 为绿色通过复合物的“顶部”进入的结合 DNA 标为紫色，白色箭头指的是结合钙离子（红色）与 RecB 核酸酶活性位点。这个结构有一缺角，以显示 DNA-ReC 与 DNA 的双链接触并将其分开，将 3′链引导到 RecB 的左面，5′尾巴结合到 RecD 的右面。（Singleton M. R. et al. 2004. Nature 432：7015，PDS Code：1W36.）图像以 PyMOL（DeLano W. L. 2002. The PyMOL molecular graphics system. 绘制，DeLano Scientific，Palo Alto，California http. www. pymol. org）

由 RecBCD 产生的单链 DNA 尾部必须被 RecA 蛋白所包被，这样才能发生重组。然而，细胞中含有一种单链结合蛋白（SSB）也可以结合到这种 DNA 上。为了保证是 RecA 而不是 SSB 结合到这些单链 DNA 尾部，RecBCD 直接与 RecA 结合以利于该蛋

白质的组装。这种结合导致了 RecB 亚基的核酸酶结构域与 RecA 蛋白质之间的蛋白质-蛋白质相互作用，并把 RecA 结合到了 DNA 的 3′端尾巴上。

Chi 位点控制 RecBCD

chi 位点能够提高约 10 倍重组发生率。这种提高在与 chi 位点直接相邻的区域最为明显。虽然 chi 位点附近 10kb 左右的区域都能观察到重组概率的提高，但这种趋势随着远离 chi 位点而逐渐降低（图 10-7）。最初，人们对这种提高只发生在 chi 位点的一侧感到迷惑，但现在已经清楚，双链断裂和 chi 位点间的 DNA 被酶切成小片段，所以不能发生重组。相反，chi 位点后的 DNA 序列遇到 RecBCD 后，保留着待重组的单链形式，并特异性地与 RecA 结合。

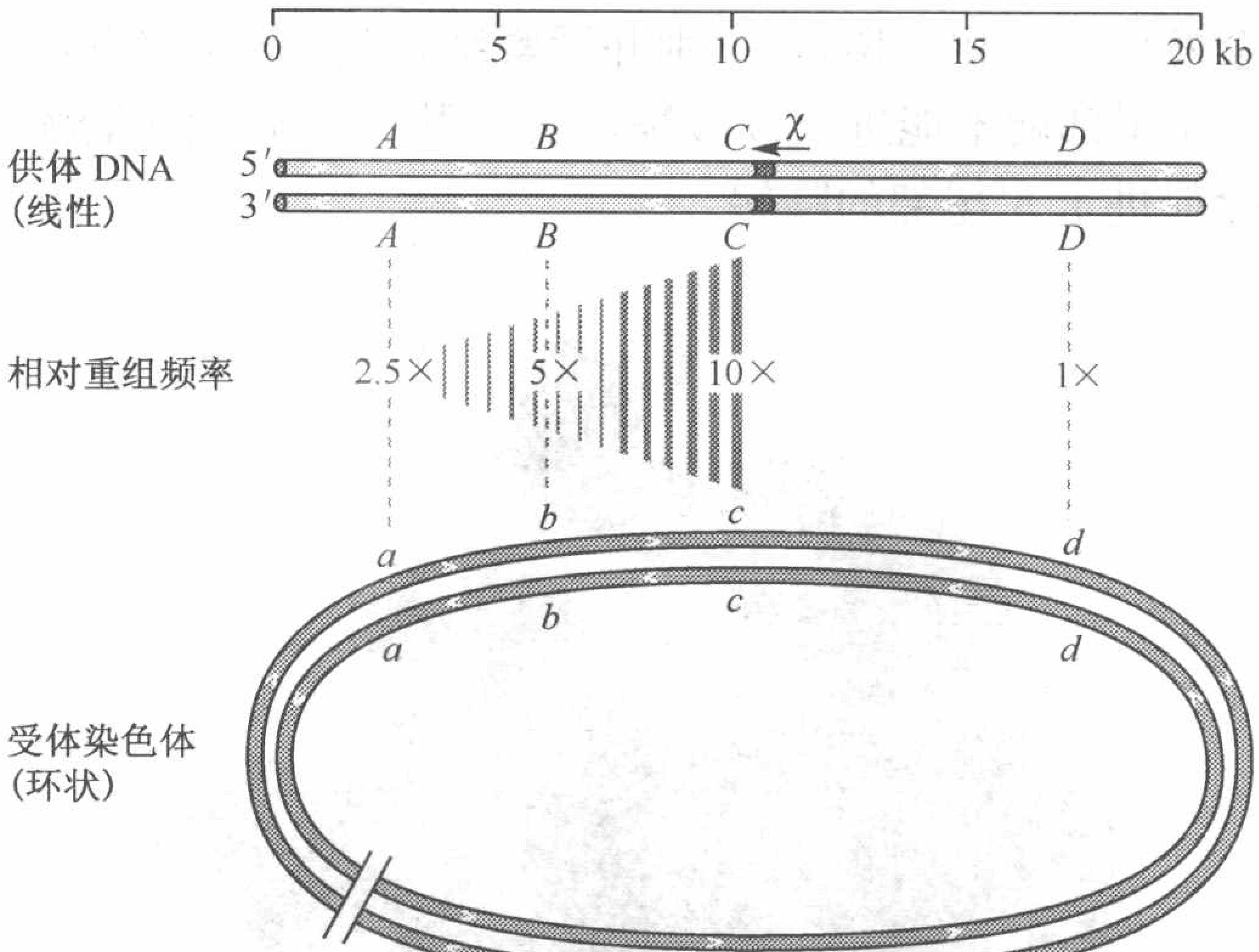


图 10-7 chi 位点的极性。本图显示了一个 chi 位点能特异性提高 chi 位点及其远处序列的重组频率。示例是一个经转导或接合进入细胞内的线性 DNA 和细菌 DNA 发生的交换事件。实际参与的 DNA 序列可能更长。例如，噬菌体转导经常传递约 80kb 的 DNA 片段。大肠杆菌染色体的长度约为 5Mb。

chi 位点控制 RecBCD 核酸酶活性的能力使之可帮助细菌细胞免受通过噬菌体介导或接合的外源 DNA 的入侵。8 个核苷酸的 chi 序列（CGTGGTGG）在大肠杆菌基因组中出现的概率极高。据预测，大肠杆菌基因组上每 65kb 才有一个，总共 80 个 chi 位点；实际上，大肠杆菌染色体序列共有 1009 个 chi 位点。由于 chi 过多出现，大肠杆菌 DNA 进入大肠杆菌细胞时易被 RecBCD 加工，产生 3′端单链 DNA，因此被激活以引发重组。反之，来自于另一株的 DNA（大肠杆菌 chi 位点未过多出现）没有过多的 chi 位点。RecBCD 作用于此 DNA 后，将导致其被过度降解，而并非被激活后而用于重组。

综而言之，RecBCD 的 DNA 降解活性有多种后果。这种降解功能对于处理 DNA 断端，使之与 RecA 组装及链侵入是必需的。在此情况下，RecBCD 促进重组。但是，RecBCD 通过降解 DNA 来激活重组，所以整个同源重组的过程中也必须有 DNA 合成以使被降解的 DNA 链再生。此外，RecBCD 的功能有时候仅仅是降解 DNA——正如降

解缺少 chi 位点的外源 DNA 一样。这时，RecBCD 可保护细胞免受引进外源序列（可能带有噬菌体或其他有害成分）可能造成的损害。

RecA 蛋白在单链 DNA 上组装并促进链侵入

RecA 蛋白是参与同源重组的关键蛋白质，是“链交换蛋白”（strand-exchange protein）酶家族的基本成员，这些蛋白质催化同源 DNA 分子的配对。配对过程涉及寻找两个分子间相配的序列和在两个分子间产生供配对的区域。

RecA 蛋白促进 DNA 配对和链交换的活性的作用可以通过使用简单的 DNA 底物进行体外实验而观察。图 10-8 列举了 DNA 配对和链交换的过程，以展示 RecA 的生化活性。这些 DNA 分子的重要特性是：① 两条 DNA 分子序列互补；② 至少其中一个分子上有使 RecA 能够组装的单链 DNA 区；③ 在互补区内存在一个 DNA 末端，使新形成的双螺旋中的 DNA 链相互缠绕。

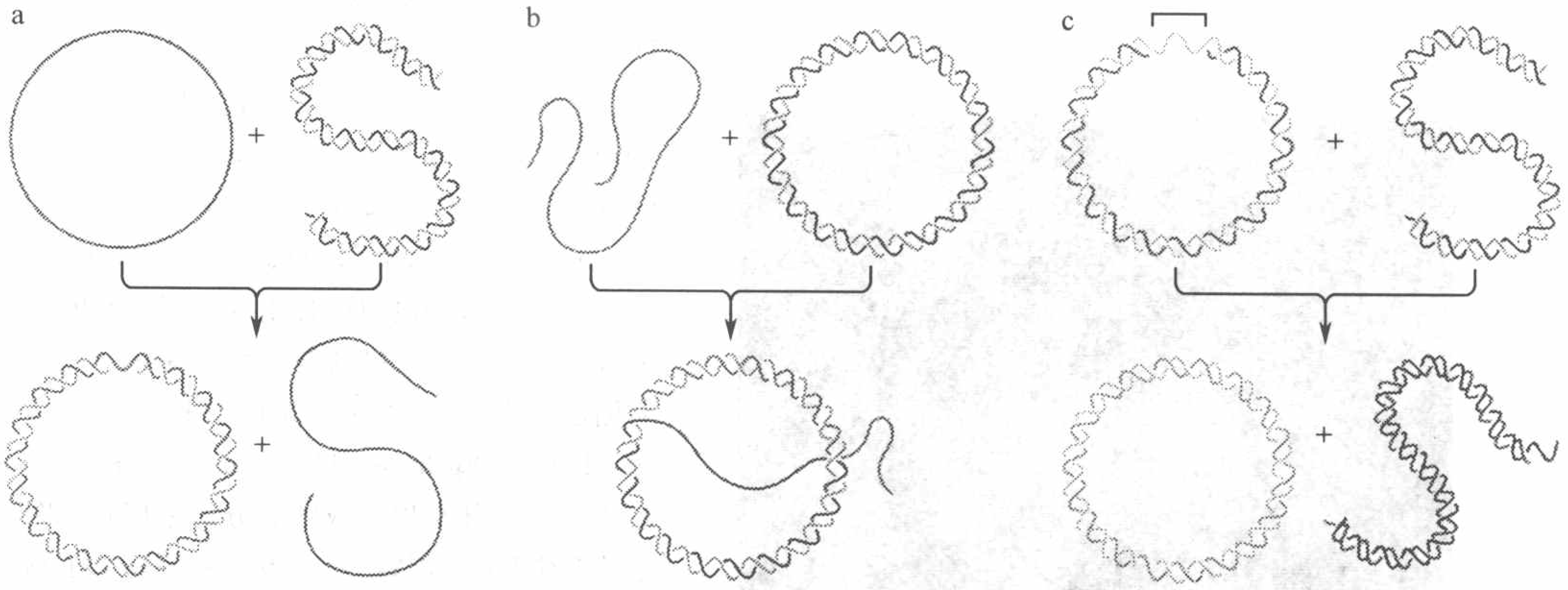


图 10-8　RecA 链交换的底物。注意 a 和 c 上的括弧显示一条 DNA 链上的缺口的位置。

具有活性的 RecA 是一种蛋白质 DNA 纤丝（图 10-9）。与多数涉及分子生物学的以小且离散的蛋白单位行使功能（如单体、双体、六聚体等）的蛋白质不同，RecA 蛋白丝则非常大，且大小不一，通常包括约 100 个 RecA 亚基与 300 个核苷酸的 DNA。RecA 蛋白丝可以容纳 1 条、2 条、3 条甚至 4 条 DNA 链。下面将会看到，有 1 条或 3 条 DNA 的 RecA 蛋白丝在重组中间体中更为常见。

此丝状结构中的 DNA 与没有蛋白质结合的单链 DNA 或标准的 B-DNA 螺旋相比高度延伸，其平均相邻碱基间距为 5Å，而不是常见的 3.4Å（第 6 章）。所以，当 DNA 和RecA蛋白结合后，DNA 分子长度增长至原来的 1.5 倍（图 10-9a）。寻找同源 DNA 序列和 DNA 链交换就是在这种 RecA 蛋白丝内进行的。

RecA 蛋白的亚基互助地结合到 DNA 上形成蛋白丝。RecA 结合与组装到单链 DNA 上的速度远快于组装到双链 DNA 上的速度，从而解释了为什么需要单链 DNA 区段作为链交换的底物。蛋白丝通过加入 RecA 亚基沿 5′→3′方向延长，这意味着以 3′端结尾的 DNA 链更易被 RecA 蛋白包裹（图 10-10）。请注意，在重组的双链断裂修复模式中，正是这种构型的 DNA 分子参与了链侵入。

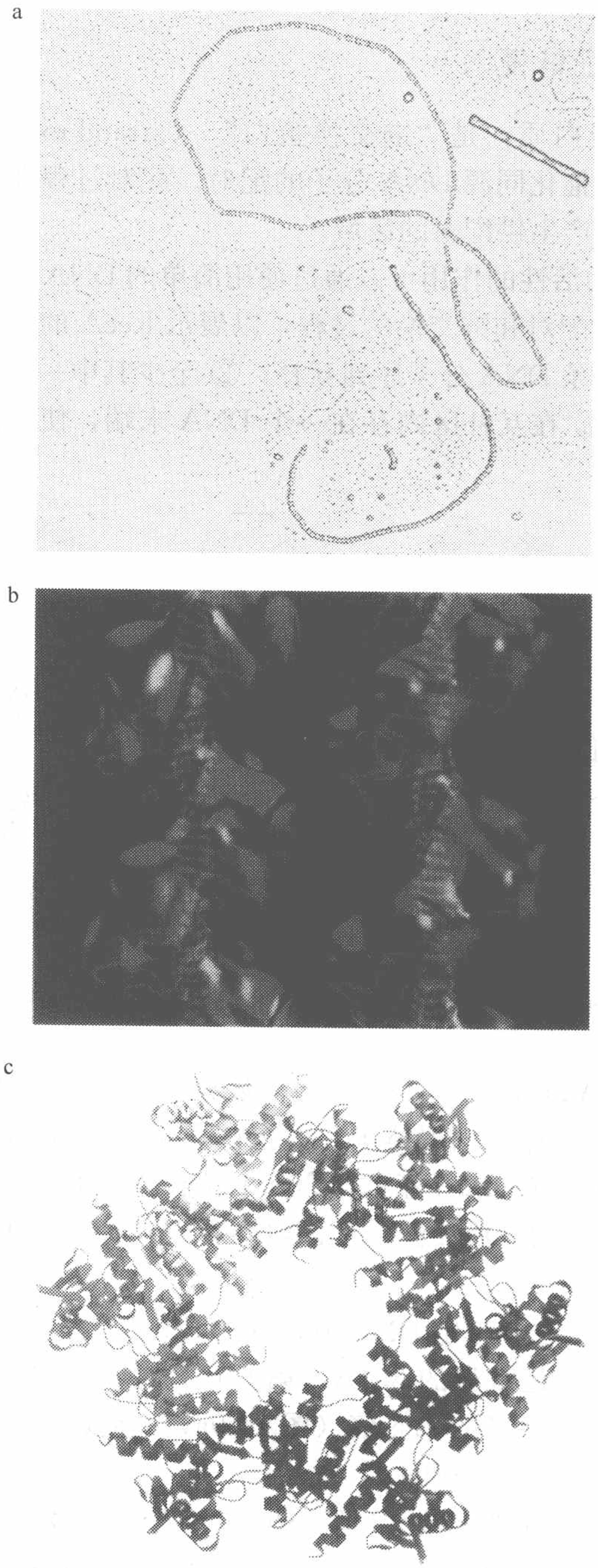


图 10-9　RecA 蛋白丝的三个视图。a. 电镜图下，全部或部分被 RecA 蛋白包裹的环状 DNA。图中有一个未被 RecA 蛋白包裹的 DNA 分子，可以看出结合 RecA 后，DNA 变长。（来源：Stasiak A. and Egelman E. H. 1988. Genetic Recombination. ed. R. Kucherlapati and G. Smith，pp. 265-307. Fig. 3.，ASM press. 3.）b. 综合多张电镜照片生成的蛋白丝的高分辨率照片。图中左侧是大肠杆菌的 RecA，右侧是酵母的链交换相关蛋白质 Rad51。（来源：Edward Egelman，University of Virginia.）c. X 射线晶体衍射生成的高分辨率照片。图示从上往下看一周螺旋丝。单个亚基标上了颜色，红色亚基靠近观察者。（Story R. M. and Steitz T. A. 1992. Nature 355：318.）镜像图用 BobScript、MolScript 和 Raster 3D 制作。

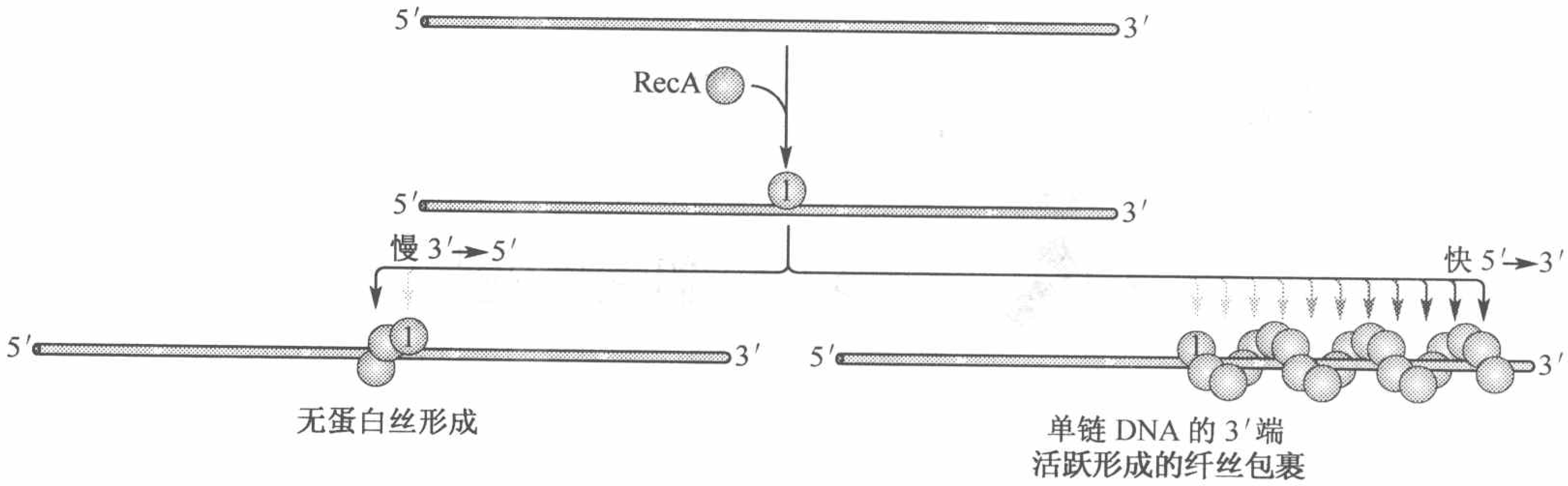


图 10-10 RecA 组装的方向性。请注意，新的 RecA 亚基在 DNA 3′侧结合到纤丝的速度远快于结合到其 5′侧的速度。因为此组装的方向性，3′延长的单链 DNA 分子更易于被 RecA 包裹；相反，5′延长的单链 DNA 难于成为蛋白丝组装的底物。

在 RecA 蛋白丝里建立新的配对 DNA

RecA 催化的链交换过程可以分成不同的反应阶段。首先，RecA 纤丝必须组装在其中一个参与的 DNA 分子上，组装发生在具有单链 DNA 区域（如单链 DNA 末端）的 DNA 分子上。这个 RecA-单链 DNA 复合体是参于寻找同源序列的活性形态。在这个过程中，RecA 必须“寻找”和蛋白丝内的 DNA 和一新的 DNA 分子之间的碱基互补。

RecA 可促进寻找同源序列，这是因为蛋白丝结构具有两个不同的 DNA 结合部位：主要位点（结合第一个 DNA 分子）与次要位点（图 10-11）。次要位点可被双链 DNA 占据。与次要位点的结合具有迅速、微弱且短暂的特性，但更重要的是此结合与 DNA 序列无关。因此，RecA 蛋白丝可以结合并快速筛选很长的 DNA 以寻找序列同源区。

RecA 蛋白丝如何检测同源序列的详细机制尚不清楚。结合在次要位点的 DNA 暂时被解链并和在主要位点的单链 DNA 做同源性比对。这种“检验”被认为是通过碱基互补配对作用进行的虽然检验初期并不破坏位于次要位点双链 DNA 的整体碱基对结构。支持这种理论的实验显示起初的配对可能涉及双螺旋数个碱基的弹出（见第 9 章的 DNA 修复过程中碱基弹出的讨论）。体外实验显示，仅 15 个碱基的序列配对就可以做为 RecA 蛋白丝发现同源区域的信号，从而启动链交换。

当一个碱基互补的区域被定位后，RecA 促进这两个 DNA 分子形成一个稳定的复合体。这个与 RecA 结合的三链结构被称为连接分子（joint molecule），通常含有数百个碱基对的杂合 DNA，实际的链交换就发生在连接分子内。在主要位点上的 DNA 链可与在次要位点上的双链 DNA 形成碱基配对，链交换需要断开链之间的一对碱基对，然后形成一组新的同样的互补碱基对。完成链交换的同时，也要求新配对的碱基链相互缠绕，形成正确的双螺旋。RecA 倾向于结合到链交换后的 DNA 产物上，正是此结合能量实际驱动交换反应，进而形成新的 DNA 构型。

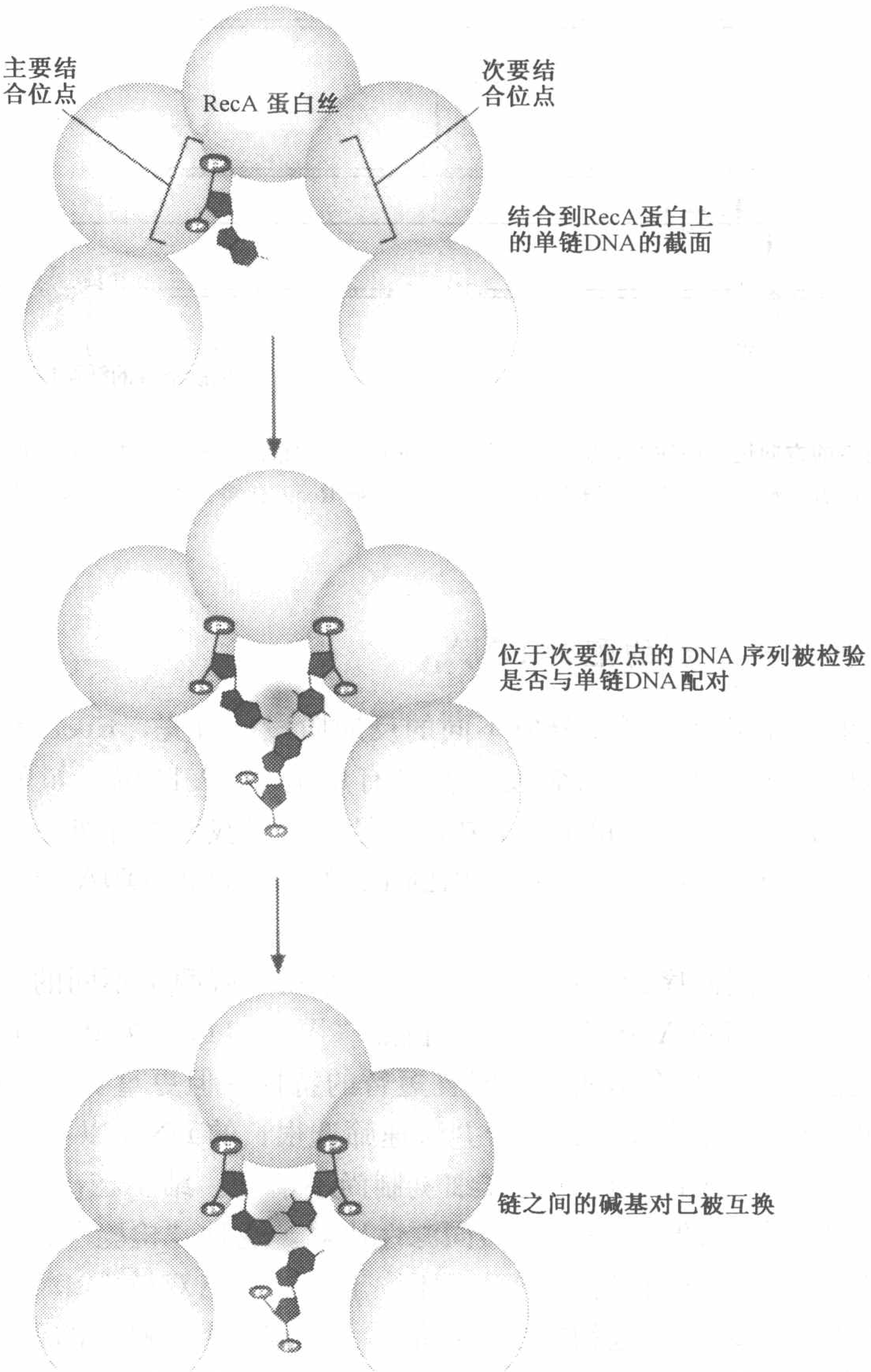


图 10-11 在 RecA 蛋白丝内寻找同源区和 DNA 链交换的两步模型。 如图 10-8c,此 RecA 蛋白丝从上往下视，新引入的 DNA 双链以蓝色表示。（来源：Howard-Flanders et al. 1984. Nature 309：215-220.）

所有的生物都有 RecA 同源物

所有的生命都有 RecA 家族的链交换蛋白质。研究得最详细的是真细菌的RecA、古细菌的 RadA、真核生物的 Rad51 和 Dmc1，及噬菌体 T4 的 UvsX 蛋白。这些蛋白质可形成与 RecA 类似（图 10-12）的蛋白丝，并以类似的方式行使相似的功能（虽然这些蛋白质的部分功能因它们特殊的细胞学角色和反应对象而发生了特化)。我们将在下面讲述 Rad51 和 Dmc1 在真核生物细胞重组中的功能。

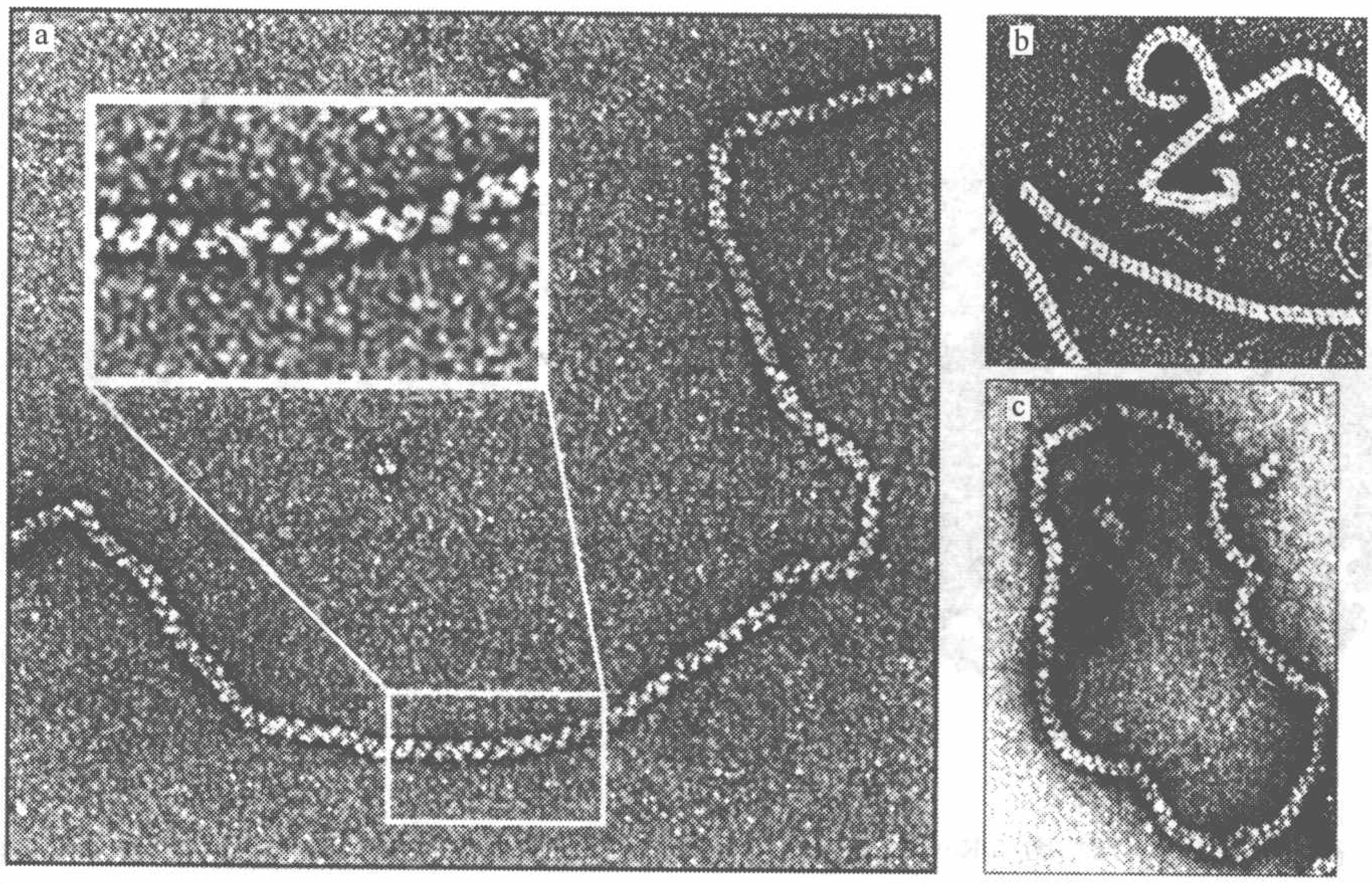


图 10-12　三类生物中类似 RecA 的蛋白质。图中显示 a. 人类 Rad51；b. 大肠杆菌 RecA；c. *Archaeoglobus. fulgidus* RadA 3 种核蛋白丝。Rad51 和 RecA 蛋白已在图 10-8 中出现过。请注意电镜图片中的局部放大图揭示的蛋白丝的类似螺旋结构。（来源：West S. C.　et al.　Nat. Rev. Mol. Cell Biol. 4：435-445.　© Macmillan.　图像由 A. Stasiak，University of Lausanne，Switzerland. 提供）

RuvAB 复合体特异性地识别 Holliday 联结体并促进分支移位

当重组的链侵入步骤完成后，两个重组的 DNA 分子被一个 DNA 分支连接起来，称为 Holliday 联结体（见前面所述）。分支点的移动需要两个同源 DNA 双螺旋交换碱基对，细胞编码的一些蛋白质可大幅提高分支移位速率。

RuvA 蛋白是一个特异性结合 Holliday 联结体的 DNA 结合蛋白，它可识别 DNA 联结体的结构，而不考虑相应的 DNA 特异性序列（见图 10-2）。RuvA 蛋白识别并结合到 Holliday 联结体上，并促使 RuvB 蛋白结合到这个位点上。RuvB 是一个六聚体 ATP 酶，类似 DNA 复制过程中的六聚体解旋酶（见第 8 章）。RuvB ATP 酶提供能量，以促进碱基对交换，从而移动 DNA 分支这种能量可以快速地在一个方向上移动 DNA 分支。一个 Holliday 联结体上的 RuvAB 复合体的结构模型（图 10-13）展示出一个四聚体 RuvA 蛋白和两个六聚体 RuvB 蛋白如何同时促进 DNA 链交换的过程。

RuvC 剪切位于 Holliday 联结体的特定 DNA 链从而结束重组

结束重组需要拆分两条重组的 DNA 分子间的 Holliday 联结体（或多个联结体）。在细菌中，RuvC 是拆分 Holliday 联结体的主要核酸内切酶。正是因 RuvC 能剪切 RecA 在体外实验中形成的 DNA 联结体而被发现并得到纯化。遗传学证据显示它的功能和 RuvA 和 RuvB 相呼应。

RuvC 引起的拆分经识别 Holliday 联结体（可能以 RuvA 和 RuvB 的复合体形式存在）后，特异地两条同一方向的同源 DNA 链打开缺口。此剪切形成了 5′端磷酸基团末

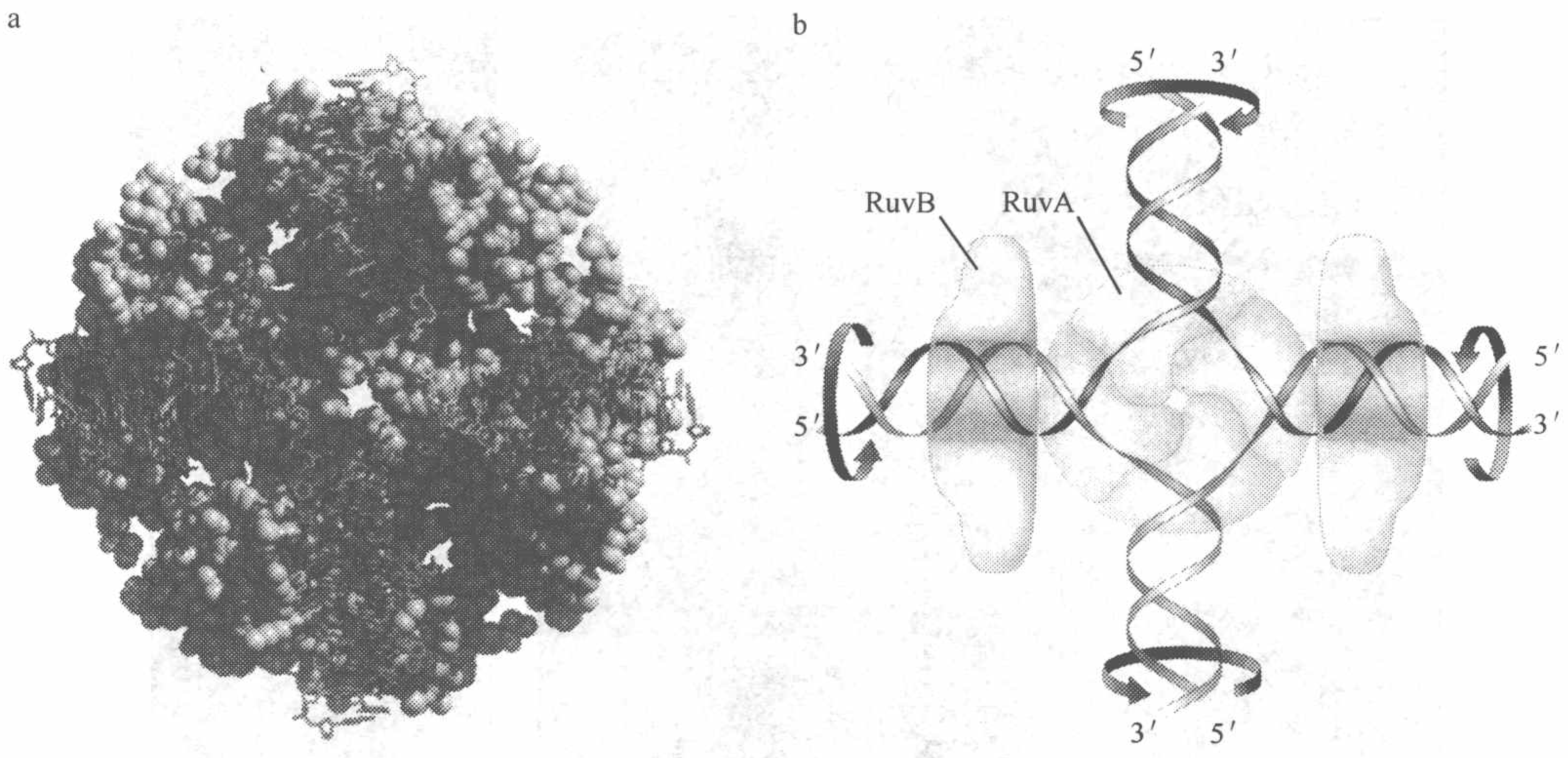


图 10-13 高分辨率 RuvA 结构图和 RuvAB 复合体与 Holliday 联结体结合的模式图。a. RuvA 四聚体的晶体结构显示该蛋白具有 4 重对称性。(Ariyoshi M.，et al. 2000. Proc. Natl. Acad. Sci. 97：8257-8262.）镜像图使用 Bobscript、Molscript 和 Raster 3D 制作。b. 有两个 RuvB 六聚体的晶体模式图。请注意，一个 RuvA 四聚体是如何以 4 重对称的形态结合到联结体上的。两个 RuvB 六聚体结合在 RuvA 对称的两侧，并起到驱动器的功能，以驱使 DNA 通过联结体。RuvB 六聚体以截面图表示，以显现穿过这些复合体的 DNA 链。(以 Yamada K. et al. 2002. Mol. Cell 10：671-681，Fig. 4重绘.)

端和3′-OH末端，这些末端可以直接被 DNA 连接酶接上。依据哪一对 DNA 链被 RuvC 剪切，连接后的重组产物可以是剪接（交换）或是补丁（未交换）的类型。图 10-14 显示了 RuvC 的结构及其如何与 DNA 联结体相互作用的模式图。

尽管 RuvC 靠识别结构而非特定序列以行使其功能，但 RuvC 剪切 DNA 还是有一定程度的序列特异性。剪切只发生在与 5′A/T-T-T-G/C 符合的一致性序列的位点，在这段序列的第二个 T 后面剪切。DNA 中这些一致性序列较多，平均每 64 个核苷酸就会有一个。这种一定程度的序列特异性可保证在联结体拆分前至少有一些分支移位发生。没有这种序列选择性，Holliday 联结体可能刚产生就被 RuvC 剪切，从而限制了参与链交换的 DNA 区域。

真核细胞的同源重组

真核细胞的同源重组具有额外的功能

如前述，在细菌中，修复 DNA 的双链断裂需要同源重组。重新启动损坏的复制叉和允许染色体 DNA 与以噬菌体感染或接合的形式进入细胞的 DNA 发生重组。同样地，在真核细胞中同源重组也是 DNA 修复和启动受损复制叉所需要的。这种

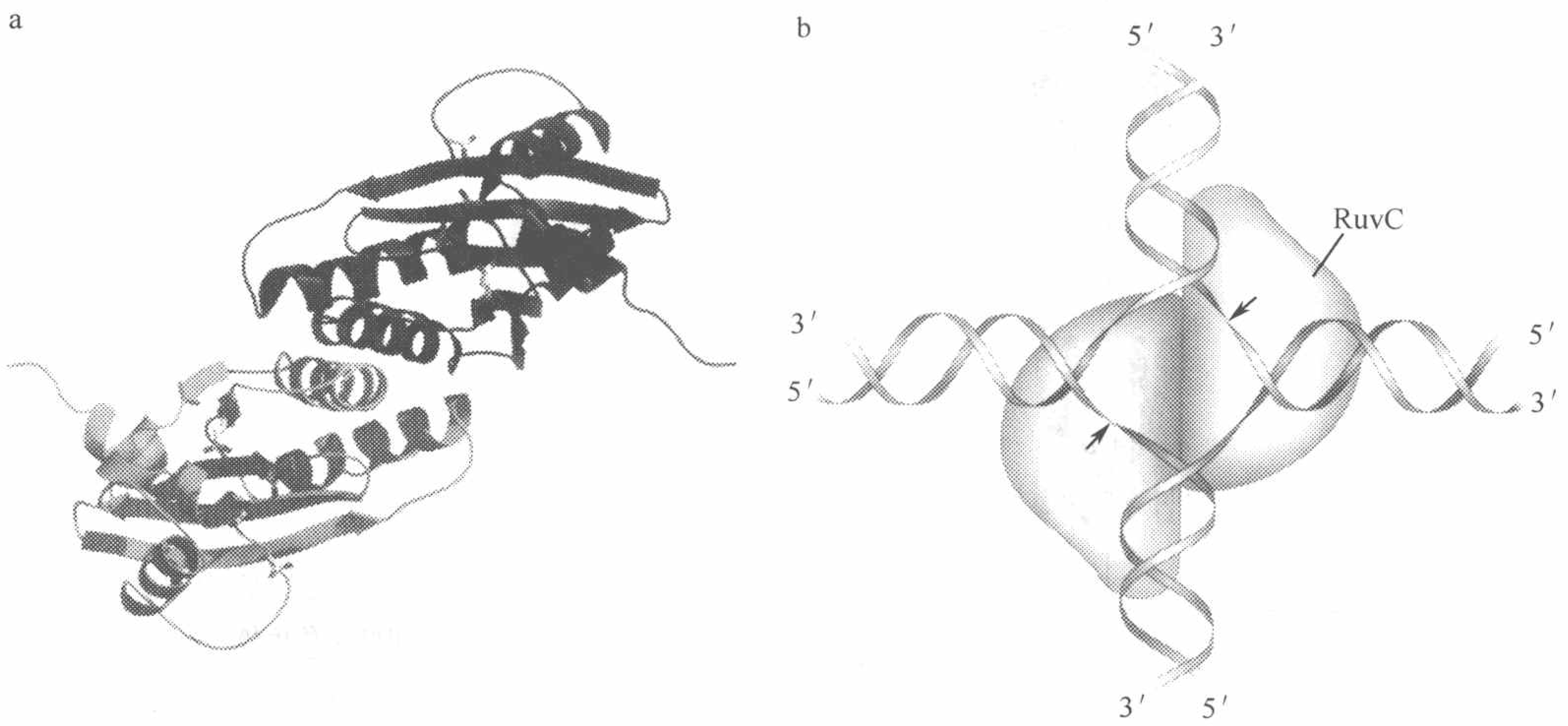


图 10-14 高分辨率的 RuvC 拆分酶结构图和与 Holliday 联结体结合的 RuvC 二聚体的模式图。 a. RuvC 蛋白的晶体结构。(Ariyoshi M., et al. 1994. Cell 78：1063-1072.) 图像使用 BobScript、MolScript 和 Raster 3D 制作。b. 一个 RuvC 二聚体结合到 Holliday 联结体上的模式图。请注意在该图中，一个 RuvC 二聚体是如何与 Holliday 联结体结合，并在两条相同的 DNA 链上引入对称的缺口。(来源：Rafferty J. B. et al. 1996. Science 274：416-421，Fig. 1b，© AAAS. 许可)

对同源重组的需要可由具重组蛋白缺陷的细胞对 DNA 破坏剂（特别是破坏 DNA 链的试剂）非常敏感所证实。此外，具有干扰同源重组突变的动物较倾向于患某种特定类型的癌症。

然而，在我们下面的讨论中可看到，同源重组在真核生物中发挥着另外的重要作用。尤其是在减数分裂的过程中，同源重组显得更为重要。在减数分裂过程中，同源重组确保染色体正确配对，从而保持基因组的完整性。这种重组可以使亲本染色体之间的基因发生重组，确保每套基因的多态性能够传给子代。

同源重组保证减数分裂中染色体的分离

正如第 7 章所述，减数分裂有两轮核分裂，减少正常双倍体细胞（2N）的 DNA 含量（2N），成为配子的 DNA 含量（1N）。图 10-15 显示了在两轮分裂中染色体是怎样分配的。分离前，细胞中的每个染色体都有两个拷贝（同源染色体），分别得自父母双方。在 S 期，这些染色体进行复制从而获得了 DNA 四倍体（4N）（图 10-16）。复制的产物——姐妹染色单体聚在一起。这时，准备进行第一次核分裂，这些复制了的同源染色体必须进行配对，并且排列在细胞的中心。正是这种同源染色体的配对需要同源重组（图 10-15）。这些事件皆需精细的时间安排。重组必须在第一次核分离前完成，从而使同源染色体正确地排列然后分开。在这个过程中，姐妹染色单体保持配对（图 10-16；图 7-16）。在第二次核分裂中，姐妹染色单体才分开。这种分裂产生 4 个配子，每一个都具有每条染色体的一个拷贝（1N 的 DNA 含量）。

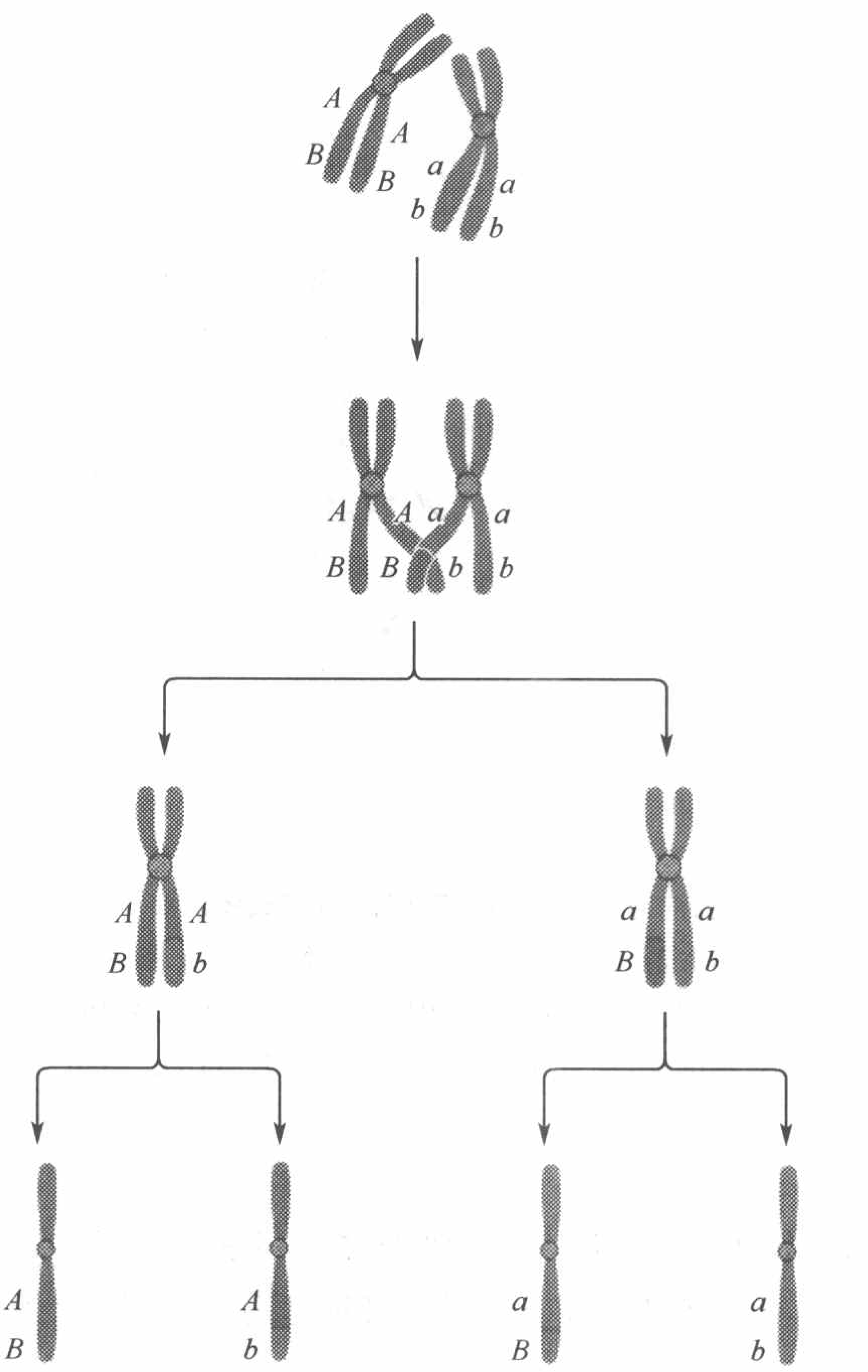


图 10-15 减数分裂的 DNA 动态过程。为了易于理解，该图只显示一对同源染色体。红色和蓝色分别显示经过一轮 DNA 复制以后的两个同源染色体。同源重组是同源染色体配对以准备第一轮核分裂所必需的。这类重组也可以引起交换，如本图中所显示的 *A* 与 *B* 基因之间。

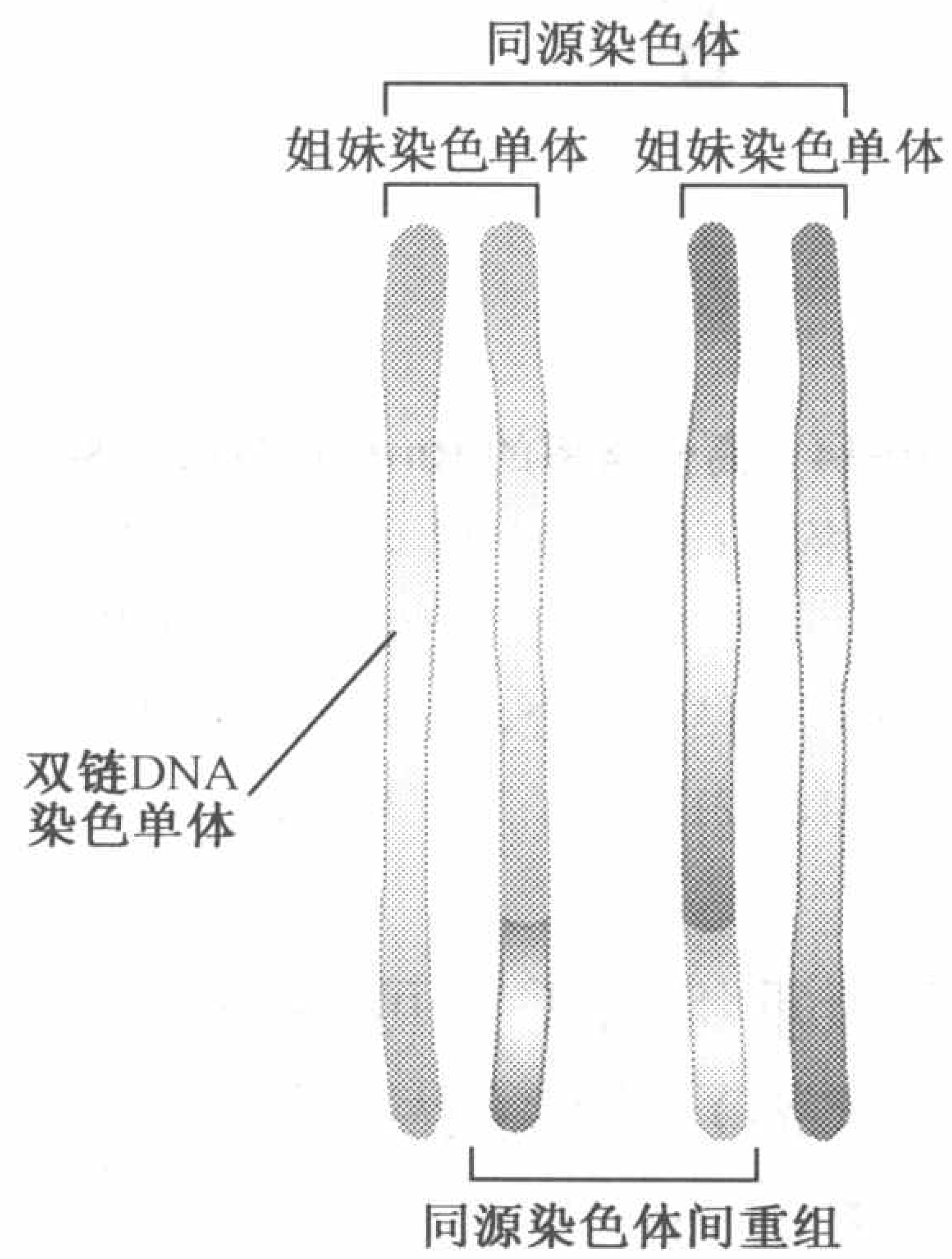


图 10-16 同源染色单体之间减数分裂重组。示已复制的双链 DNA 分子即染色单体，配对的称为姐妹染色单体，发生在非姐妹染色单体对之间的重组是由 Dmc1 介导的。（见图 10-19）

如果没有重组，染色体通常无法在第一次减数分裂过程中正确地排列，从而极易引起染色体的丢失。染色体的不恰当分离，被称之为不分离（non disjunction），会导致大量配子没有正确的染色体组成。具有过多或过少染色体数量的配子在受精后都不能正常发育，因而，失败的同源重组往往表现于低生殖能力。这种在减数分裂过程中发生的同源重组被称为减数分裂重组（meiotic recombination）。

减数分裂重组也常常引起两个同源亲代染色体间基因的交换，这种遗传交换（如图 10-16 显示的那样）可以用细胞学方法观察到（图 10-17，左侧图示）。一个重要的结果是亲本 DNA 分子上的等位基因在下一代发生了重新分配。

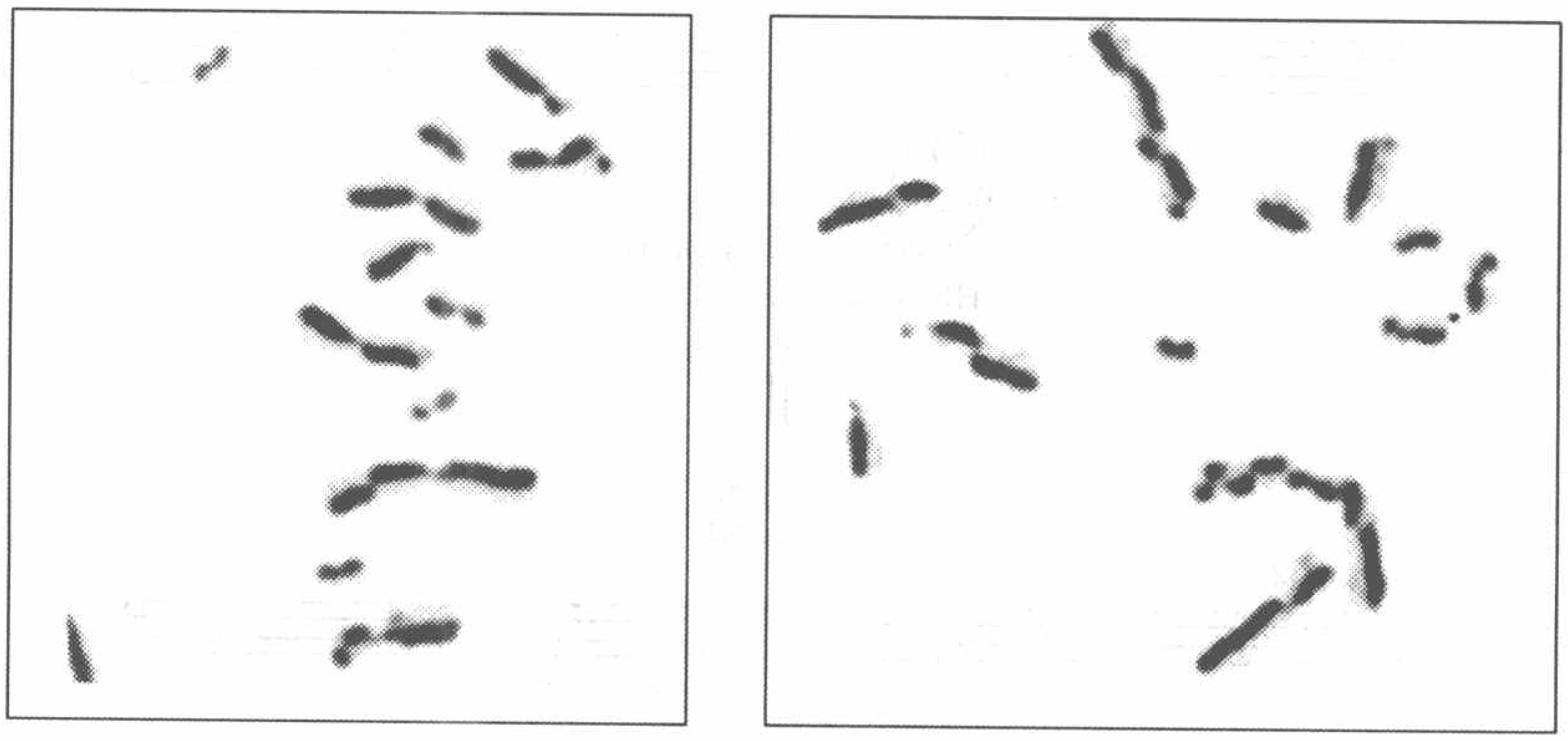


图 10-17 交换的细胞学图示。仓鼠组织培养细胞中直接观察到的染色单体相互交换。经过 Giema 染色的染色体，DNA 双链中胸腺嘧啶均被溴脱氧尿苷替代并被染浅色，如果只有一条链被替代，则呈深色。经过两个世代生长于溴脱氧鸟苷中，经复制的新的染色单体仅有一条链被替代，而另一姐妹染色单体则双链均被替代。如此，姐妹染色单体就可以区分。交换则可以很容易的以浅深交替带而显现出来（左侧）。同样，如果用 DNA 破坏试剂处理正处于有丝分裂的细胞，也能看到相间的重组染色体（右侧）。（Sheldon Wolff and Jody Bodycote 惠赠）

减数分裂中程序化生成的 DNA 双链断裂

细胞成功地完成减数分裂所需要的发育程序包括启动很多一般生长过程中所不需要的基因表达。其中的一个就是 *SPO11*，该基因编码一个蛋白质，可引入染色体 DNA 的双链断裂进而启动减数分裂重组。

Spo11 蛋白可以在很多染色体位置上切断 DNA，它对序列没什么选择性，但却必须发生在减数分裂的一特定时期。Spo11 参与的 DNA 切割正好发生在已复制的同源染色体开始进行配对的时候。Spo11 切割的位置尽管普遍，但这些切割位点并非随机地分布在 DNA 上，相反，它们多分布在核小体包装疏松的染色体区域，如控制基因转录的启动子区域（见第 7 章和第 17 章）。DNA 出现较高概率的双链断裂区域，也会发生高频率的重组。如此，常被 Spo11 切割的 DNA 位点，如 chi 位点，也是重组的热点。

DNA 断裂的机制如下：Spo11 蛋白中一个特异的酪氨酸侧链可攻击磷酸二酯骨架，从而切断 DNA，并在蛋白质与受损的 DNA 链间生成共价复合物（图 10-18），Spo11 的两个亚基在两个 DNA 链的相差两个核苷酸处切开，形成一个交错的双链断裂。Spo11 与 DNA 拓扑异构酶及位点特异性重组酶具有同样的 DNA 剪切机制（见第 6 章和第 11 章）。实际上，蛋白质序列比较显示 Spo11 和这些酶都有一定的同源性。

Spo11 的切割涉及一个蛋白质-DNA 共价复合物，这导致了两个结果。其一，Spo11 切割处 5′端的 DNA 可共价连接该酶。正是这些与 Spo11 相连的 5′DNA 端形成了单链 DNA 尾端的起始点，这种尾端可引起类似 RecA 的蛋白质的组装，并启动 DNA 链侵入（如下文所述）。第二，被切断的 DNA 磷酸二酯键的能量储存在蛋白质-DNA 的连接中，这样 DNA 链能通过切割反应简单逆转（化学机制见图 11-5）而被重接。这种重接发生在细胞接到停止减数分裂的信息之后。

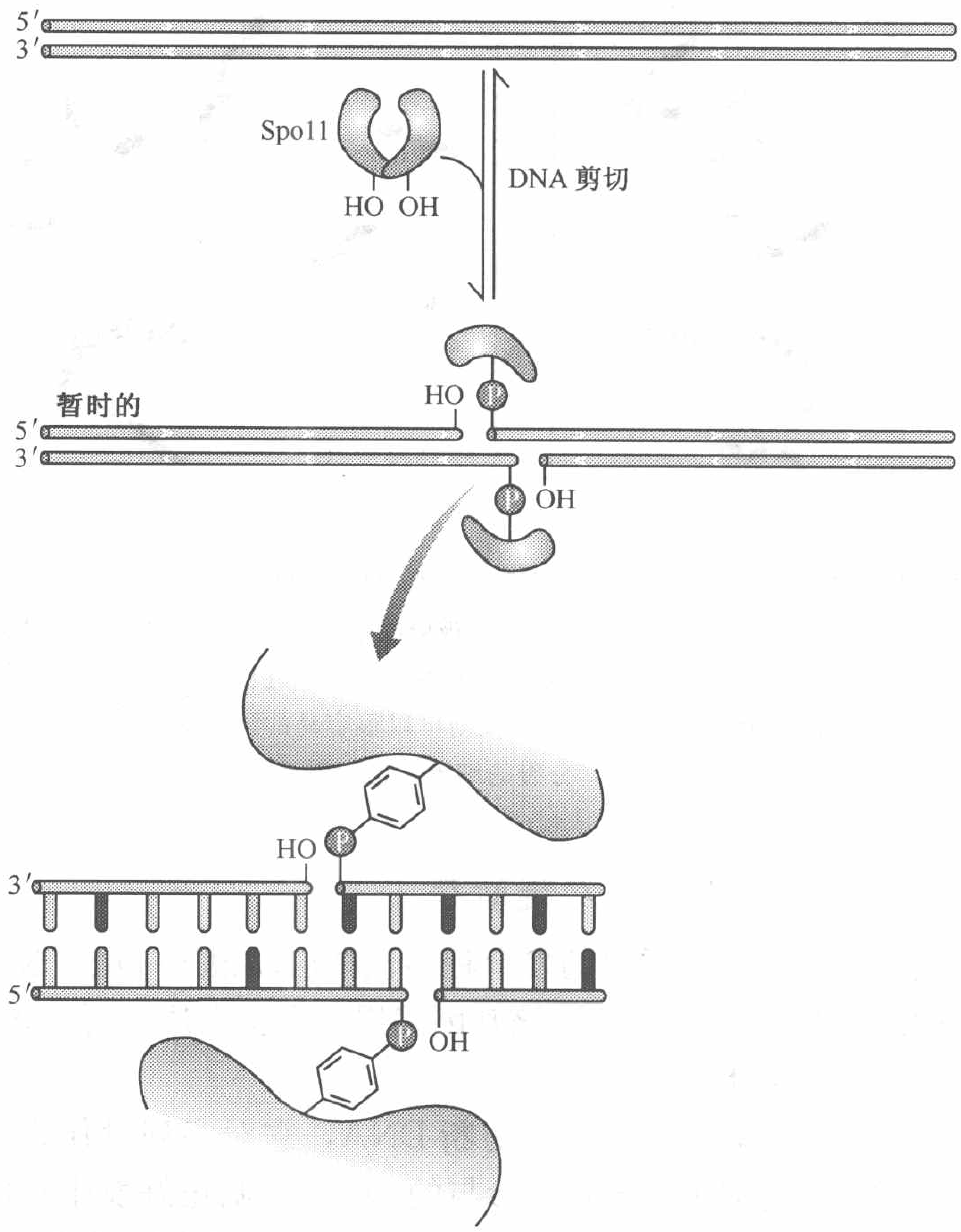


图 10-18 Spo11 的切割机制。Spo11 蛋白中酪氨酸的—OH 基团对 DNA 进行攻击，生成一个蛋白质-DNA 共价连接。Spo11 的两个亚基是生成双链 DNA 断裂所必需的，一个亚基攻击其中一条 DNA 链。注意，由于这样的切割机制，DSB 能够通过切割反应的简单逆转而重接。

MRX 蛋白作用于被切开的 DNA 末端，以组装类 RecA 的链交换蛋白

Spo11 催化的 DNA 双链断裂生成单链区域，该区域用于组装类 RecA 的链交换蛋白。如细菌中观察的 RecBCD 路径，这个处理过程产生了很长的 3′端单链 DNA 片段（图 10-19）。在减数分裂重组中，MRX 酶复合物负责催化该 DNA 处理过程。MRX 酶复合物尽管与 RecBCD 不同源，但也是一个多亚基的 DNA 核酸酶。MRX 酶复合物是由 Mre11、Rad50 和 Xrs2 三个亚基组成，并以三个亚基的首字母命名。

断裂处的 DNA 处理完全作用于 5′端的 DNA 链上，即与 Spo11 蛋白共价结合的 DNA 链（如前述）。3′端的 DNA 链不被降解，所以此 DNA 处理反应称之为 5′→3′端切除。MRX 正是通过 5′→3′的反应，从而生成长具 3′端的单链 DNA，通常可达 1kb 或更长。MRX 酶复合物还被认为能去除与 DNA 相连的 Spo11。

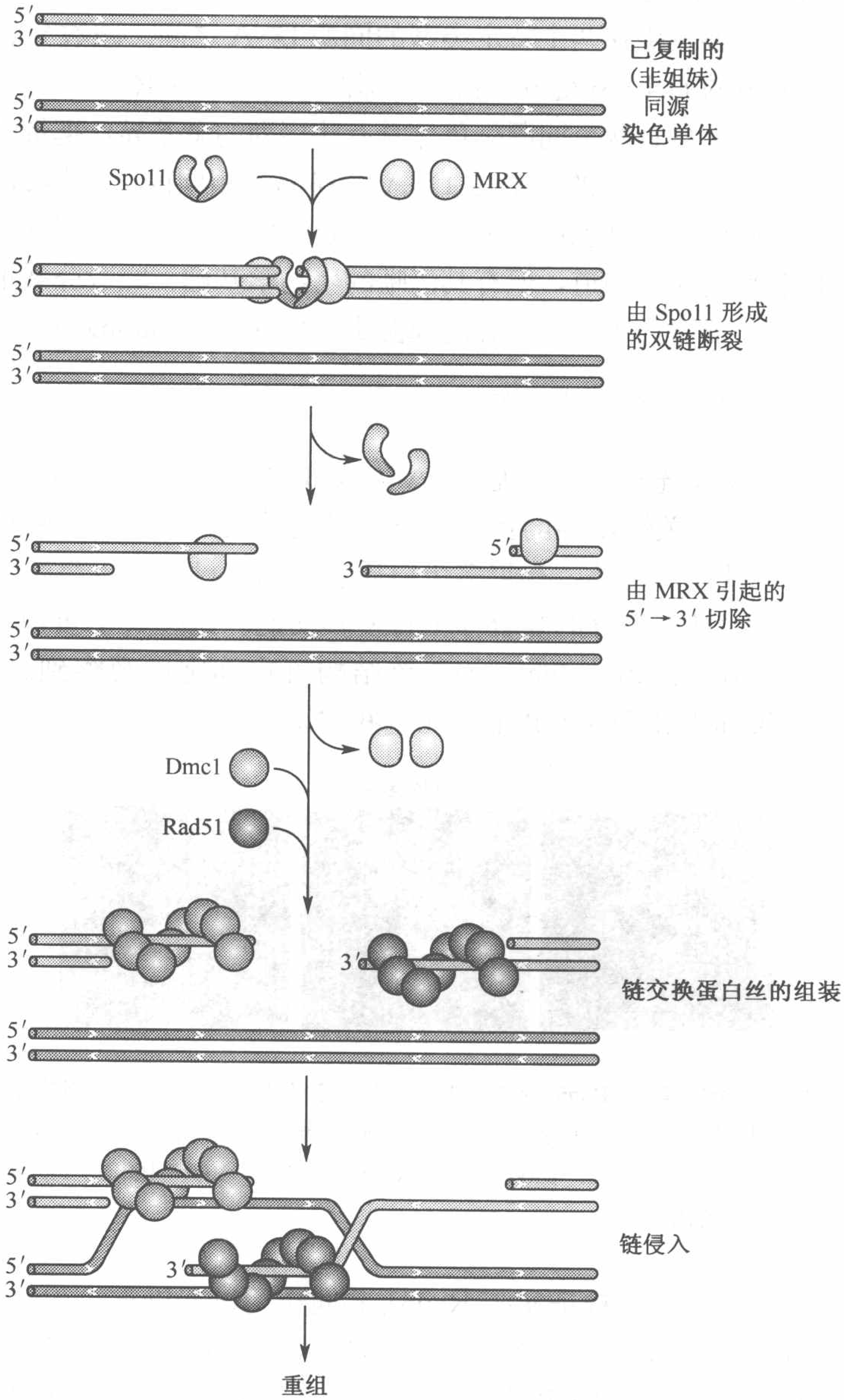


图 10-19 减数分裂重组的途径。减数分裂过程中双链断裂的形成需要 Spo11 和 MRX 复合体的参与。观察指出，DSB 的形成和后续的链处理，通常是由多种蛋白质的协调作用的偶联而完成。MRX 蛋白质负责切除断裂位点的 5′端 DNA，然后链交换蛋白 Dmc1 和 Rad51 组装到单链 DNA 尾上。以上两个蛋白质都参与了重组，但还不清楚它们如何共同作用。为便于理解，本图中显示它们正在形成独立的蛋白丝。(经 Lichten M. 2001. Curr Biol. 11：R253-R256，Fig. 2. © Elsevier. 许可重绘)

Dmc1 是专一在减数分裂重组中行使功能的类 RecA 蛋白

真核生物编码两种被充分阐明的、与细菌 RecA 蛋白同源的蛋白质：Rad51 和 Dmc1。这两种蛋白质都在减数分裂重组中发挥作用。尽管 Rad51 广泛地表达在进行细

胞有丝分裂和减数分裂的细胞中，Dmc1 则仅在细胞进入减数分裂时被表达。

减数分裂中的链交换发生在两个特定类型的同源 DNA 之间。减数分裂重组发生在每个染色体都有 4 套完整的 DNA 双链分子拷贝，即两个同源染色体各复制生成了一对姐妹染色单体时（图 10-16）。尽管两个同源染色体具有些序列的微差异，携带不同的基因不同的等位基因，但这 4 套染色体中的大部分 DNA 序列都是相同的。有趣的是，依赖 Dmc1 的重组倾向发生于非姐妹同源染色单体之间，而不是姐妹染色单体之间（图 10-16）。尽管这种选择机制的机制目前还不知道，但有其清晰的生物学基础：减数分裂重组可通过促进同源染色体间的交联，而有助于等待分裂的同源染色体联会（alignment）。

多种蛋白质共同促进减数分裂重组

如前述，参与形成双链断裂、处理 DNA 生成 3′端结尾的单链 DNA 和参与减数分裂重组过程中的链交换等关键步骤的蛋白质已被发现并定性。遗传实验显示很多其他的蛋白质也参与了这一过程。此外，有很多蛋白质表现出与已知的重组酶间的作用，并且它们似乎是在一个很大且多组分的复合体中行使其功能。这些大型蛋白质-DNA 复合体被称为重组工厂（recombiantion factory），该结构可在细胞中观察到。例如，图10-20 显示在减数分裂中 Rad51 和 Dmc1 共定位在这些“工厂”里。

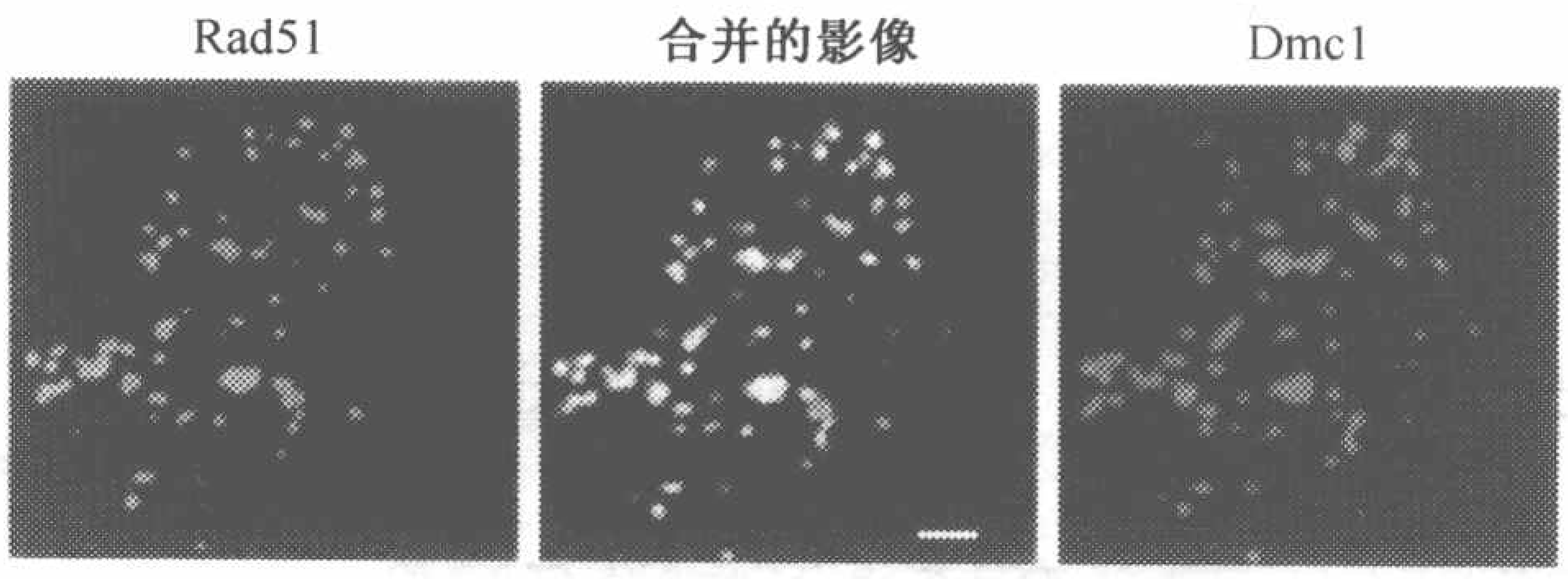


图 10-20 在细胞减数分裂过程中，Rad51 和 Dmc1 蛋白共同定位于“重组工厂”。用荧光标记 Rad51（绿色）和 Dmc1（红色）抗体，以免疫染色法检测蛋白质。两种蛋白质在一起的时候，合并的影像呈黄色。（来源：经 Shinohara M. et al. 2000. Proc. Natl. Acad. Sci. 97：10814-10819，Fig. 1A，© National Acodemy of Sciences 许可使用）

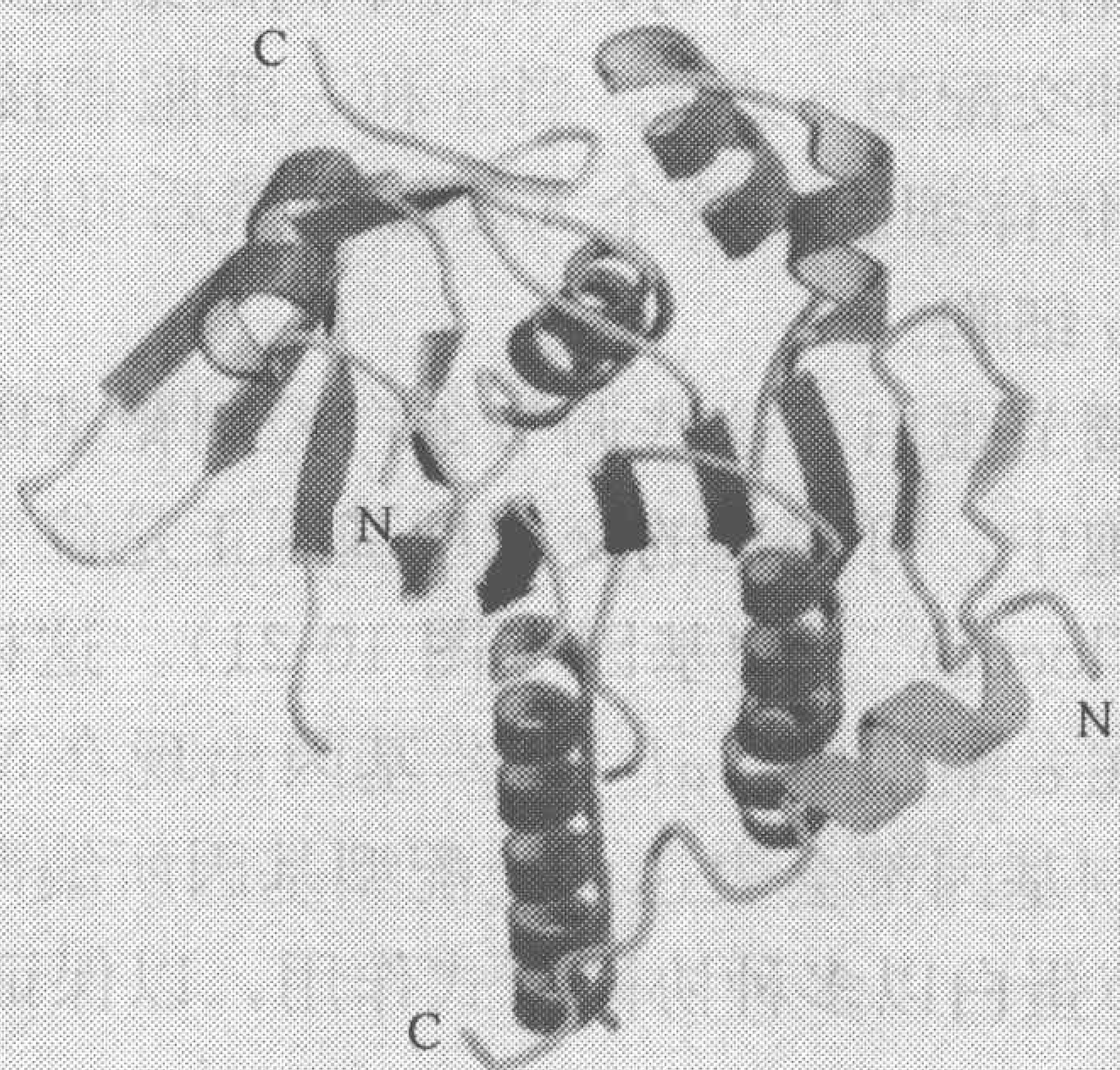
还有很多其他种类的蛋白被证实与 Rad51 共同作用以促进重组和双链 DNA 断裂的修复。Rad52 是另外一个与 Rad51 相互作用的必要重组蛋白。Rad52 启动 Rad51 DNA 蛋白丝（Rad51 的活性形态）的组装。Rad52 通过对 RPA（真核细胞中主要的单链 DNA 结合蛋白）的拮抗作用而启动 Rad51 蛋白丝的组装。因此，Rad52 与已知大肠杆菌的 RecBCD 蛋白有相似的活性，RecBCD 蛋白可帮助 RecA 结合在原本被 SSB 所结合的单链 DNA 上。

Rad52 蛋白也能促进互补单链 DNA 分子的退火和碱基配对，这种活性在重组起始阶段的双链配对反应中也发挥着重要的作用。BRCA2 基因的产物也参与 Rad51 介导的双链断裂修复（见框 10-2）。

医学链接

框 10-2 肿瘤抑制基因 BRCA2 的产物和 Rad51 蛋白相互作用并调控基因组的稳定性

BRCA2 基因对于维持基因组的稳定性具有重要的作用。在人类中，近一半的家族性乳腺癌都被认为是由于 BRCA2 基因的突变造成的。这种癌症的易感性至少部分来自 BRCA2 在 Rad51 介导的双链断裂修复中的直接作用。当细胞被培养在对 DNA 有损伤的化学试剂中时，可以很明显地观察到 Rad51 聚集体首先被组装起来以激活 DNA 修复。BRCA2 基因缺陷的细胞不能在细胞受损时在细胞核中组装起这些聚集体同时也不能进行相应的 DNA 修复。BRCA2 蛋白和 Rad51 蛋白直接相互作用（见框 10-2 图 1），它们的相互作用可能对 Rad51 蛋白在细胞中正确定位以利于 DNA 修复以及对调节 Rad51 蛋白的活性非常重要。因此，与 BRCA2 基因突变相关的显著表现型说明，在真核生物中包括在人类中，DNA 双链断裂修复和 Rad51 依赖性的同源重组是至关重要的。



框 10-2 图 1 Rad51 蛋白和 BRCA2 蛋白重复基序相互作用复合体的三维结构。 生化与结构研究表明，BRCA2 上被称作 BRC 基序的含有保守重复序列的特异区段是和 Rad51 蛋白作用的重要位点。其中的一个重复基序 BRC4 被证实与 Rad51 蛋白具有很高的亲和性。结构分析更精确地揭示了 BRC4 是如何与 Rad51 蛋白形成复合体的。在本图中，Rad51 蛋白的 α 螺旋用紫色显示，β 折叠用蓝色显示，BRC 肽段的重复序列用绿色显示。每条序列都标注了氨基和羧基末端。（来源：经 Pellegrini L. et al. 2002. Nature 420: 287-293, Fig. 1a. © Macmillan. 许可使用）

通过与细菌的模拟比较，我们推测真核细胞也编码启动重组过程中分支移位和 Holliday 联结体拆分的蛋白质。事实上，具有启动这些功能的酶已经被发现了。例如，Rad51C（一种含有类 Rad51 蛋白的复合体）和另一种蛋白 XRCC3 已经被证实具有 Holliday 联结体拆分酶的活性。

如我们所看到的，减数分裂重组可使同源染色体联会，并促进它们之间的遗传交换。这些重组反应通常导致亲本染色体间的交换。然而，依赖重组中间物的 Holliday 联结体的拆分结果，通过双链断裂修复途径的重组也能够产生非交换产物（见上文）。这些事件可为成功的减数分裂提供必需的染色体配对，但未在染色体遗传结构上留下可

检测的变化。

然而，即便是非交换重组也会产生遗传后果，如导致基因转变（gene conversion）。基因转变发生在当一个基因的等位基因丢失并被其他等位基因所替代之时。关于基因转变是如何发生在有丝分裂生长期细胞和减数分裂细胞中的例子，将在下面介绍。

交配型的转换

除了促进 DNA 配对、DNA 修复和遗传交换外，同源重组还能够改变染色体特定位置的 DNA 序列。这种类型的重组有时是为了调控基因的表达。例如，经由转换不同交配型基因出现在该生物基因组中可被表达的特定位置上，重组可对酿酒酵母的交配类型进行控制。

酿酒酵母是种单细胞真核生物，并可以 3 种不同的细胞类型生存（见第 22 章）。酿酒酵母的单倍体细胞有两种交配型 a 或 α。当 a 和 α 细胞相接近时，它们能够进行融合（即交配）形成一个 a/α 双倍体细胞。这个 a/α 细胞然后可以经过减数分裂形成两个单倍体 a 细胞和两个单倍体 α 细胞。

交配型基因编码转录调节蛋白。这些调节蛋白控制靶基因的表达，该靶基因的产物决定细胞类型。在一个细胞中，可表达的交配型基因正是位于该细胞基因交配型基因座（mating-type locus，MATlocus）上的基因（图 10-21）。这样，在 a 细胞中 *a1* 基因出现在 *MAT* 基因座上，而在 α 细胞中，*α1* 和 *α2* 基因出现在 *MAT* 基因座上。在双倍体细胞中，此两种交配型控制基因都会表达。交配型基因所编码的调节蛋白，一起与其他 3 种细胞类型中发现的调节蛋白以多种协同方式作用，以保证在不同细胞类型中被正确表达（见第 17 章）。

如我们正在讨论的，细胞能够通过重组转换它们的交配型。各细胞除了 a 或 α 基因存在于 *MAT* 基因座上以外，另外还有一套 a 和 α 基因存在于基因组的其他位置（但并不表达）。这些额外的沉默拷贝所在的基因座称之为 *HMR* 和 *HML*（图 10-21）。

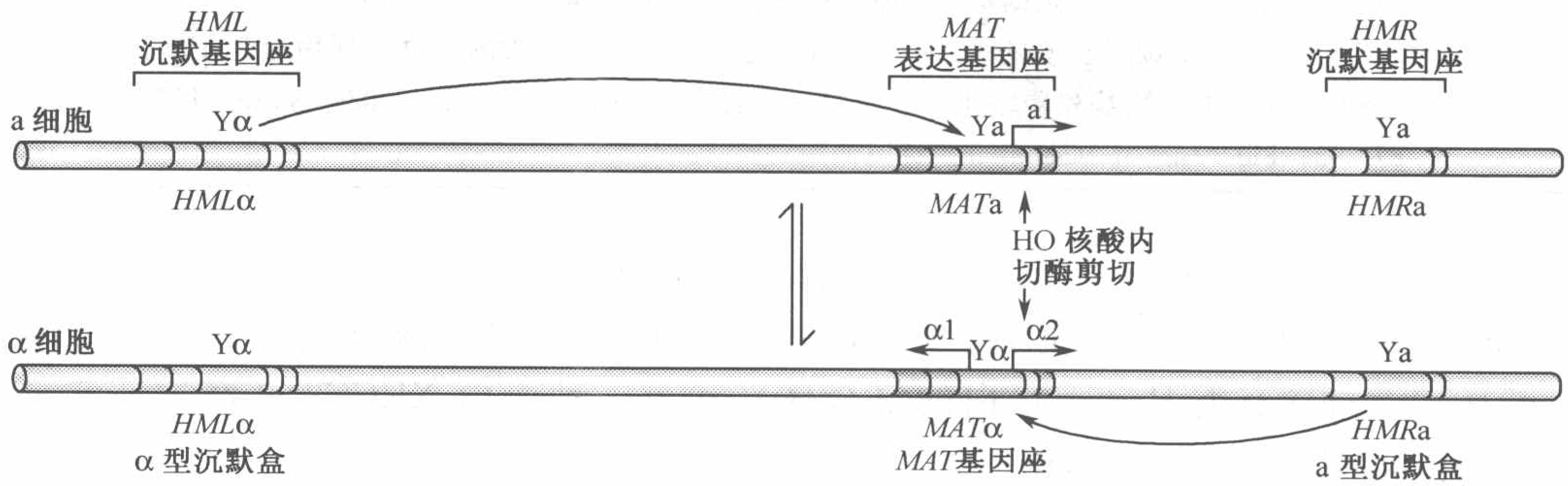


图 10-21　编码交配型信息的遗传基因座。尽管 3 号染色体带有 3 个交配型基因座，但是只有在 *MAT* 基因座上的基因可以表达。*HML* 编码一个沉默拷贝的 α 基因，而 *HMR* 编码一个沉默拷贝的 a 基因。如果重组发生在 *MAT* 和 *HML* 位置上，a 细胞就会转换成 α 细胞。如果重组发生在 *MAT* 和 *HMR* 位置上，α 细胞转换成 a 细胞。（来源：经 Haber J. E. 1998. Annu. Rev. Genet. 32：561-599，Fig. 3. © Annual Reviews 许可修改）

所以，这些 *HMR* 和 *HML* 基因座被称为沉默盒（silent cassettes）。它们的功能是提供一个遗传信息的"储藏室"，用来改变细胞交配型。这种转换需要通过同源重组使遗传信息从 *HM* 基因座转换到 *MAT* 基因座。

特定位点的双链 DNA 断裂引发交配型转换

交配型转换是由 *MAT* 基因座的双链断裂引发的，这个反应由特异的 DNA 切割酶——HO 内切核酸酶（HO endonuclease）来完成。HO 基因的表达被严格地控制，以确保转换在适当时候发生。这个调控机制将在第 17 章中进行介绍。HO 是一种序列特异性的内切核酸酶。酵母染色体上唯一携带 HO 识别序列的是交配型基因座。HO 在染色体上切割形成一个黏性的断裂，与 Spo11 切割相比，HO 只是水解 DNA，并不与 DNA 断裂处共价相连。

HO 引起的断裂，其 5′→3′的 DNA 切除反应与减数分裂重组的机制一样。所以，切除依赖于 MRX 蛋白复合体，并对带 5′端的 DNA 链具有特异性，而带 3′端 DNA 链保持稳定，而不被核酸酶消化。一旦生成长的 3′单链 DNA 尾巴，它们就被 Rad51 和 Rad52 蛋白包裹（和其他帮助组装重组蛋白-DNA 复合体的蛋白质）。这些被 Rad51 蛋白包裹的 DNA 链寻找同源染色体区域以启动链的侵入和遗传交换。

交配型的转换是单向性的，即序列信息（尽管并不是真正的 DNA 片段）从 *HMR* 和 *HML* "移动"到 *MAT* 基因座，但是绝对不会从反方向进行。这样，在重组过程中，被切掉的 *MAT* 基因座总是接受者，而 *HMR* 和 *HML* 基因座在整个重组过程中是不变的。这种方向性源于 HO 内切核酸酶不能切割其在 *HML* 和 *HMR* 基因座上的识别序列，因为此区域的染色质结构使得酶无法接近这些位点。

由 Rad51 包裹位于 *MAT* 基因座的 3′单链 DNA 末端可"选择" *HMR* 或 *HML* 基因座的 DNA 进行链侵入。如果 *MAT* 基因座的 DNA 序列是 a，则侵入会发生在有 α 序列的"储备"拷贝的 *HMLR*；反之，如果 *MAT* 带有 α 基因，则侵入会发生在储备有 a 序列的 *HMR*。重组之后，被选择的 *HM* 基因座的信息也出现在 *MAT* 基因座。这种遗传改变不是从 *MAT* 到 *HR* 基因座的信息相互交换。这种非交互的重组事件是基因转变的一个特例。

交配型转换是一种基因转变事件，与交换无关

尽管 DSB 修复途径可以解释交配型转换的重组机制，但是大量的证据显示，在链侵入步骤之后，此重组途径不同于 DSB 修复机制。其中一个提示是在交配型转换的过程中从未发现属于交换的重组产物。回顾在 DSB 修复途径中，产生两种 Holliday 联结体中间型的拆分产物：剪接（交换产物）和补丁（非交换产物）（图 10-3）。根据 DSB 修复模式，这两种产物发生的概率相似，但是，在交配型转换中从未观察到有交换产物存在。所以，无 Holliday 联结体中间体的重组模型更适于解释交配型转换。

为解释没有发生交换事件的基因转变的机制，提出了一种新的重组模型，称之为**依赖合成的链退火**（synthesis-dependent strand annealing，SDSA）。图 10-22 显示在这种机制下，如何发生交配型转换。如上所述，首先在重组位点引入 DSB（图 10-22a），经过链侵入之后（图 10-22b），侵入的 3′端作为引物启动新的 DNA 合成（图 10-22c、d）。值得注意

的是，与 DSB 修复途径所不同的是，一个完整的复制叉在该处形成。前导链和后随链可同时进行 DNA 复制，但是与平常的 DNA 复制不同，新合成的链自模板上置换下来，其结果是新的双链 DNA 片段被合成，再连接到最初被 HO 内切酶切断的 DNA 位点上，并被 MRX 切除。这个新的片段具有与模板相同的 DNA 序列（图 10-22 中的 *HMRa*）。

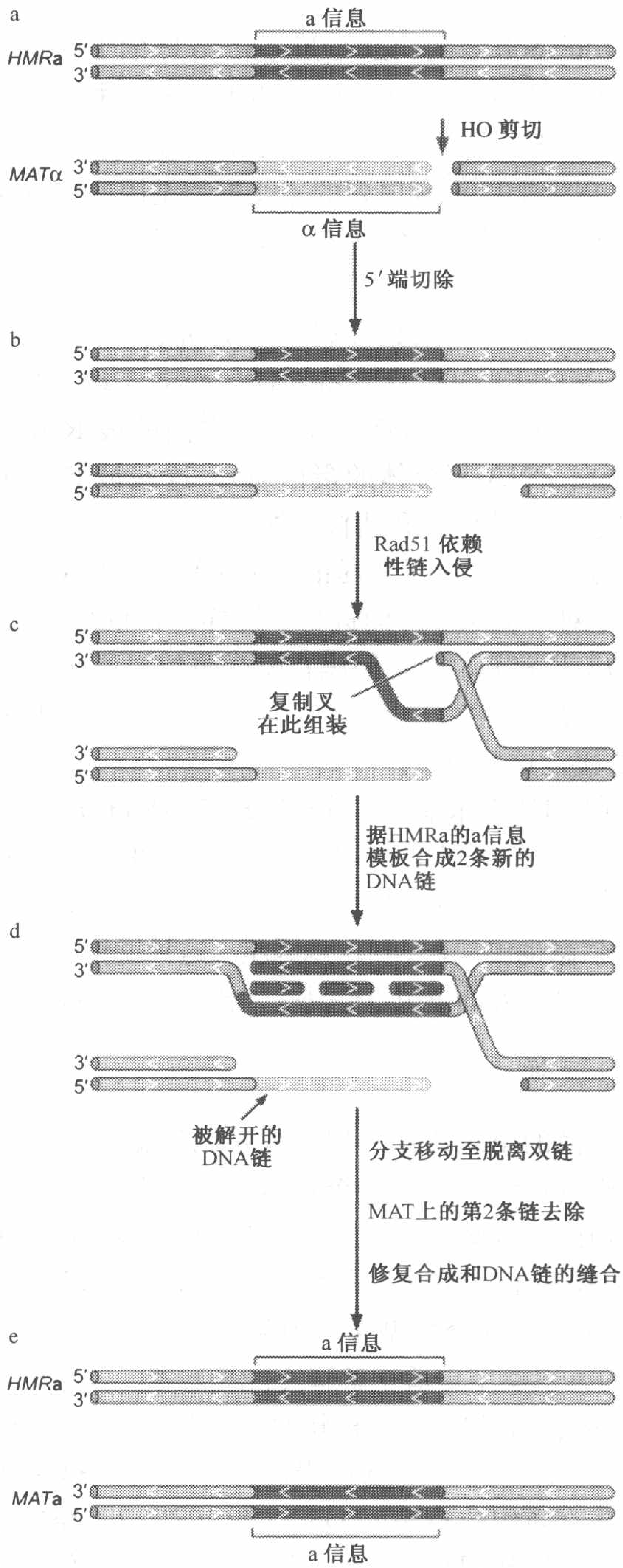


图 10-22 交配型转换的重组模型：依赖合成的链退火（synthesis-dependent strand annealing，SDSA）。图示位于 *MAT* 基因座基因转变的步骤。*HMR* 和 *MAT* 区域显示为绿色；编码 a 信息的 *HMR* 为深绿色，编码 α 信息的 *MAT* 呈橙绿色。当 SDSA 结束时，起初位于 MAT 区域的 α 信息被替换，即转变成为 HMR 区域的 a 信息。

完成重组需要切除在 *MAT* 基因座上的另一条 DNA 旧链（MRX 不会切除具 3′端的链）（图 10-22d 下面一条链）。然后，新合成的 DNA——其互补 DNA 分子信息的精确拷贝，替代原有的 DNA 信息。这种机制更恰当地解释了不需要切除 Holliday 联结体的基因转变。通过这个模型，交配型重组时没有发现交换产物不再是一个谜。

同源重组机制的遗传结果

如在本章开头所讨论的，最初的同源重组机制模型在很大程度上是为了解释该过程的遗传结果。现在重组的基本步骤已被了解，我们有必要回顾一下同源重组的过程及其如何改变 DNA 分子而产生特异的遗传变化。

同源重组的一个重要特点是，无论是何种序列，只要有足够的相似区域，它就可以发生在任何两个 DNA 区域间。现在我们知道了原因所在，同源重组没有一个步骤需要特异 DNA 序列的识别。至于部分依赖序列的步骤（如经 chi 位点转化的 RecBCD 和被 RuvC 蛋白所剪切的 DNA），其所识别的序列则很常见。两个 DNA 分子间发生重组的关键步骤在于 RecA 家族的一个链交换蛋白质能成功地将两条分子配对，而任何具有能形成正确碱基配对的正常 DNA 链都可以指导这个过程。

重组大致上独立于序列的事实导致一个推论，即任何两个基因间的重组频率和这些基因间的距离成正比。因为 DNA 区域可被平等地用来启动一个成功的重组事件，所以可观察到此正比性。同源重组的这个基本特性使我们能够用重组率来制作一个遗传图，以显示染色体上基因的顺序和距离。

当 DNA 的一个区域在形成重组时不具有这种“平均”概率时，将遗传图和物理图比较就会产生曲解（图 10-23）。高重组概率的区域称为“热点”，而重组概率低的区域称为“冷点”。因此，中间有热点的两个基因，在遗传图上要比在物理图上的同一区域相距更远。反之，中间有冷点的基因在遗传图上的距离比它们的物理图距离要更接近。

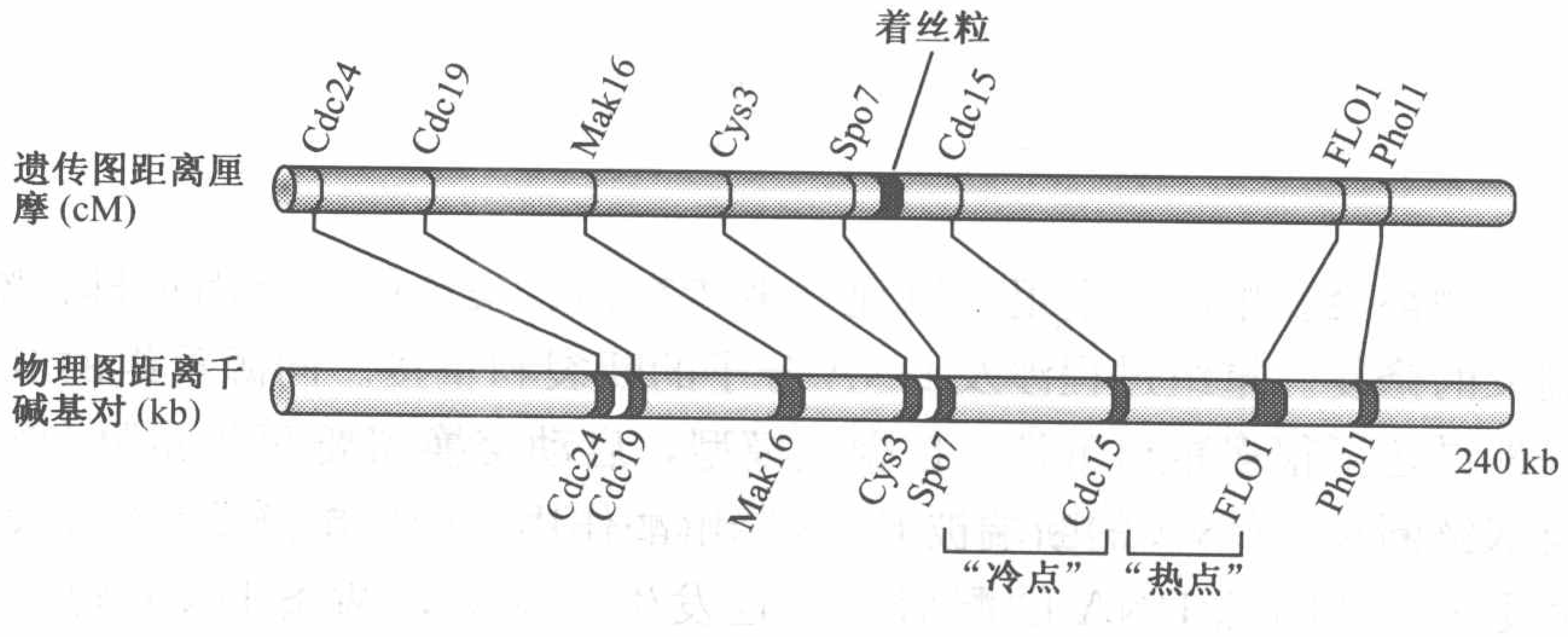


图 10-23 酵母染色体一段典型区域的遗传图和物理图的比较。标记显示了不同基因的位置。注意遗传图上 Spo7 和 Cdc15 之间的区域，因为低交换频率，遗传图小于物理图；相反，Cdc15 和 FL01 之间的区域，因为高频率交换，遗传图大于物理图。（来源：经 Alberts B. et al. 2002. Molecular biology of the cell，4th edition，p. 1138，Fig. 20-14. © Garland Science/Taylor & Francis LLC. 许可改绘）

我们已遇到两个能够从分子水平上解释染色体上存在冷点和热点的例子。靠近 chi 位点和 Spo11 剪切位点的区域是热点，拥有高于平均的启动重组概率，而这类位点较少的区域则是冷点。

重组时的 DNA 修复基因转变的起因之一

同源重组的另一个遗传结果是基因转变。我们曾在酵母交配型转换的重组特例中介绍了基因转变的概念。然而，基因转变也常发生在正常同源重组事件中，如引起细菌中遗传交换和减数分裂中染色体的配对。

为了演示减数分裂重组中的基因转变，假设一个减数分裂期的细胞在其同源染色体上分别为等位基因 *A* 和 *a*。DNA 复制后，将产生 4 个该基因的拷贝，基因型分别是 *A*、*A*、*a* 和 *a*。如果没有基因转变，将产生 2 个带有 *A* 等位基因的配子和 2 个带有 *a* 等位基因的配子。如果所产生的配子的基因型是 *A*、*a*、*a*、*a*（或者是 *A*、*A*、*A*、*a*），则说明发生了一个将 *A* 置换成 *a*（或相反）的基因转变事件。这是怎么发生的呢？

在双链断裂修复过程中，基因转变以两种方式发生。首先，考虑如果 *A* 基因与双链断裂处很接近时将会发生什么。这时当 3′单链 DNA 末端侵入同源双螺旋链并延伸，则可能会复制 *a* 信息，这个过程可以导致替换重组后染色体上的 *A* 信息（图 10-4d）。

基因转变的第二种机制涉及修复重组中间体所产生的碱基错配。假如链侵入或分支移位包括 *A*/*a* 基因，则可能会形成一段异源双链 DNA，在一条链上有 *A* 序列而另一条链上有 *a* 序列（图 10-24 和图 10-2d 的放大图）。带有碱基错配的 DNA 区域能被细胞错配修复酶识别和作用（我们在第 9 章讨论过），这些酶能特异地修复 DNA 上的碱基错配。当碱基错配酶发现碱基错配后，会在两条链中的一条上切除一小段 DNA。然后，一个具修复作用的 DNA 聚合酶填补这个缺口，并产生正确的配对序列。当作用于重组中间体时，错配修复酶将随机选择被修复的一条链。因此，修复后产生的双链带有编码 *A* 信息的序列还是编码 *a* 信息的序列，取决于哪条链被修复酶修复，这就产生了基因转变。

小结

所有的生物都会发生同源重组，以允许遗传交换、染色体上基因重排、断裂 DNA 及停滞复制叉的修复。重组过程涉及 DNA 分子的断裂和重接。同源重组的双链断裂修复途径很好地描述了很多重组事件。依据此模型，启动交换需要两条同源 DNA 分子中的一条具有双链断裂。DNA 的断端被 DNA 降解酶作用，产生单链 DNA 片段。这些单链 DNA 片段参与其同源 DNA 的配对。一旦发生了配对，两条 DNA 分子就被称为 Holliday 联结体的分支结构所连接。在 Holliday 联结体处切断 DNA，以拆分联结体并结束重组。Holliday 联结体可被两种方式剪切：其中一种方法会产生交换产物，即两个亲本 DNA 分子被共价连接；另一种剪切联结体的方法不产生交换产物而仅是形成一个重组 DNA 的“补丁”。

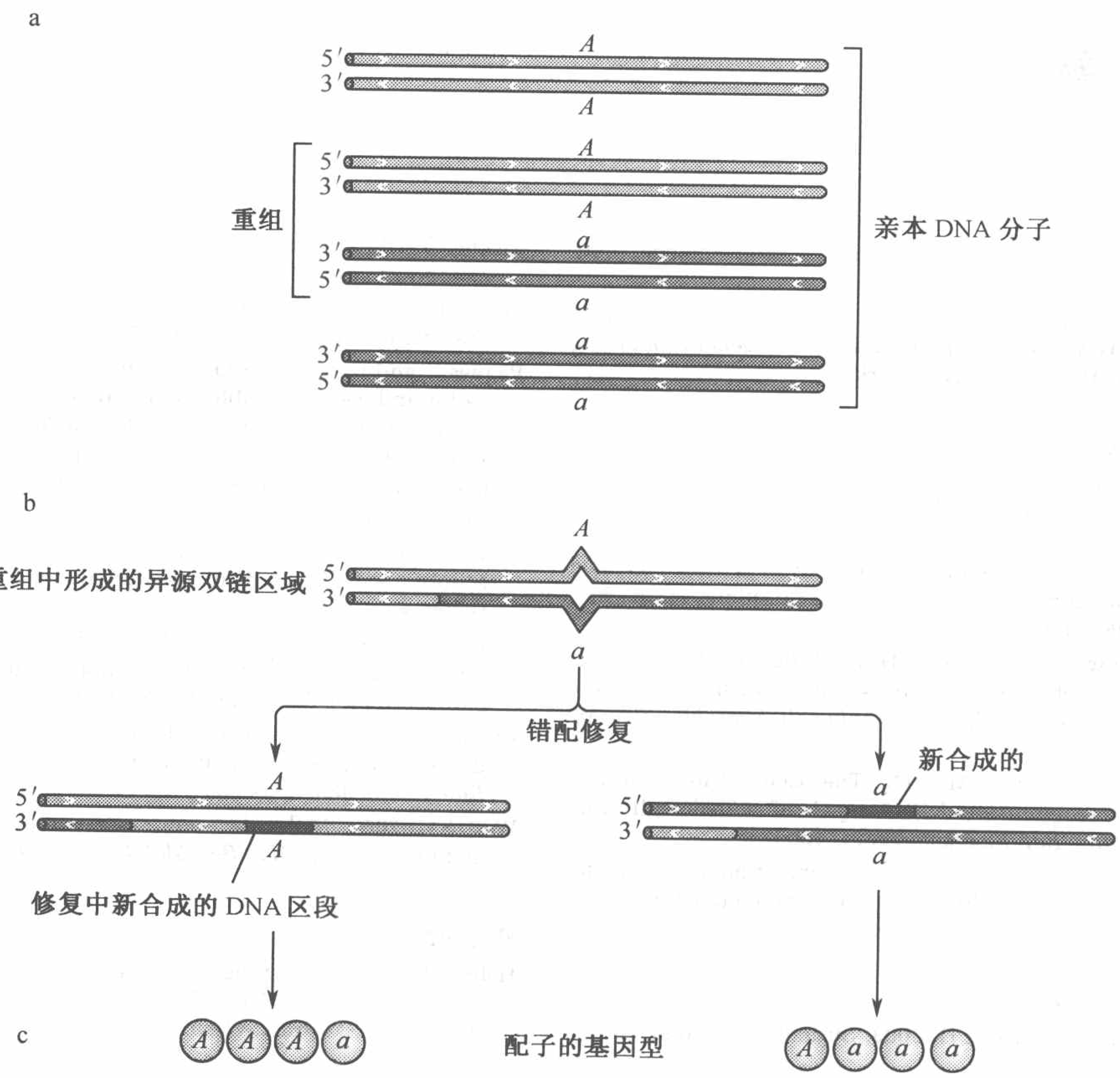


图 10-24　错配修复重组中间体的异源双链 DNA 可以导致基因转变。

细胞编码催化所有同源重组步骤的酶。关键酶是链交换蛋白。其中，大肠杆菌的 RecA 是个先例；类 RecA 的蛋白质存在于所有的生命体中。类 RecA 的链交换蛋白促进两条 DNA 分子寻找同源序列和重组中间体的链交换。RecA 以一个大的 DNA-蛋白质复合体（RecA 蛋白丝）发挥功能。真核生物的细胞编码两个链交换蛋白，分别为 Rad51 和 Dmc1。其他重要的重组酶是产生 DNA 双链断裂并启动重组的 DNA 剪切酶。这些蛋白质（Spo11 和 HO）只在真核细胞中出现。作用于 DNA 断端而产生所需要的单链末端的核酸酶，分别是原核生物的 RecBCD 酶和真核生物中的 MRX 酶复合体。此外，还有促进移动（分支移位）和剪切（拆分）Holliday 联结体的其他酶。

减数分裂中，重组对第一次核分裂前同源染色体的正确配对是必不可少的。因此，重组被高度调控以保证发生在所有的染色体上。Spo11 DNA 切割酶和 Dmc1 链交换蛋白都在重组反应中起特异性的作用。同源重组有时也用来调控基因表达。酵母的交配型转换是这种调控的绝佳例证，同时也是一个基因转变的例子。已经有一个被称为依赖合成的链退火（SDSA）的新模型描述一些同源重组事件，该模型是在分析交配型转换的机制的过程中产生的。这种机制可以产生基因转变型的遗传交换产物但不会导致交换的发生。

（赵辉　译　王敦梅　胡松年　刘韧　杨焕明　校）

参考文献

书

Brown T.A. 2007. *Genomes*, 3rd ed. Garland Science, New York.

Griffiths A.J.F., Miller J.H., Suzuki D.T., Lewontin R.C., and Gelbart W.M. 2000. *An introduction to genetic analysis*, 7th ed. W.H. Freeman, New York.

细菌的重组

Court D.L., Sawitzke J.A., and Thomason L.C. 2002. Genetic engineering using homologous recombination. *Annu. Rev. Genet.* **36:** 361–388.

Cox M.M. 2001. Recombinational DNA repair of damaged replication forks in *Escherichia coli:* Questions. *Annu. Rev. Genet.* **35:** 53–82.

Kowalczykowski S.C., Dixon D.A., Eggleston A.K., Lauder S.D., and Rehrauer W.M. 1994. Biochemistry of homologous recombination in *Escherichia coli. Microbiol. Rev.* **58:** 401–465.

Lusetti S.L. and Cox M.M. 2002. The bacterial RecA protein and the recombinatorial DNA repair of stalled replication forks. *Annu. Rev. Biochem.* **71:** 71–100.

Smith G.R. 2001. Homologous recombination near and far from DNA breaks: Alternative roles and contrasting views. *Annu. Rev. Genet.* **35:** 243–274.

真核细胞的重组

Eichler E.E. and Sankoff D. 2003. Structural dynamics of eukaryotic chromosome evolution. *Science* **301:** 793–797.

Keeney S. 2001. Mechanism and control of meiotic recombination initiation. *Curr. Top. Dev. Biol.* **52:** 1–53.

Page S.L. and Hawley R.S. 2003. Chromosome choreography: The meiotic ballet. *Science* **301:** 785–789.

Pâques F. and Haber J.E. 1999. Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae. Microbiol. Mol. Biol. Rev.* **63:** 349–404.

Pastink A., Eeken J.C., and Lohman P.H. 2001. Genomic integrity and the repair of double-strand DNA breaks. *Mutat. Res.* **480-481:** 37–50.

Prado F., Cortes-Ledesma F., Huertas P., and Aguilera A. 2003. Mitotic recombination in *Saccharomyces cerevisiae. Curr. Genet.* **42:** 185–198.

Symington L.S. 2002. Role of *RAD52* epistasis group genes in homologous recombination and double-strand break repair. *Microbiol. Mol. Biol. Rev.* **66:** 630–670.

van den Bosch M., Lohman P.H., and Pastink A. 2002. DNA double-strand break repair by homologous recombination. *Biol. Chem.* **383:** 873–892.

West S.C. 2003. Molecular views of recombination proteins and their control. *Nat. Rev. Mol. Cell Biol.* **4:** 435–445.

酵母交配型的转换

Haber J.E. 2002. Switching of *Saccharomyces cerevisiae* mating-type genes. In *Mobile DNA II* (ed. N.L. Craig et al.), pp. 927–952. ASM Press, Washington, D.C.

第11章 位点特异性重组和DNA转座

本章概要

- 保守性位点特异性重组（CSSR）
- 位点特异性重组的生物学作用
- 转座
- 转座子及其调控的实例
- V(D)J 重组

DNA 是一种非常稳定的分子。通过前面章节的学习，我们了解到：DNA 的复制、修复以及同源重组过程都保持了很高的保真度，为一个物种的基因组在传代过程中保持几乎完全一致提供了保证。然而，重要的是还有一些遗传过程造成了 DNA 序列的重排，从而导致了一个更具动态性的基因组结构，这些过程就是本章的主题。

很多重要的 DNA 重排都是由两类重要的遗传重组造成的：保守性位点特异性重组（conservative site-specific recombination，CSSR）和转座重组（一般称作转座）（transpositional recombination）。CSSR 发生在两段特定的序列之间（图 11-1）。相反，转座则是发生在特定序列和非特定 DNA 位点之间。由这些重组活动促成的生物学过程包括：病毒感染过程中病毒基因组插入到宿主细胞 DNA、DNA 片段倒位造成基因结构的变化，以及转座因子（transposable element）（又叫“跳跃”基因）从一个染色体位点到另一个位点移动。

这些 DNA 重排对染色体结构和功能有着深远的影响。对许多物种而言，转座是自然突变的主要来源，并且人类基因组中几乎一半的序列来自转座子（尽管目前大多数元件都不具活性）。而且，正如我们将看到的，病毒感染和脊椎动物免疫系统的发育都严重地依赖这种特殊的 DNA 重组。

保守重组和转座具有共同的关键机制。已知的蛋白重组酶（recombinase）能够识别 DNA 分子上要发生重组的特定序列，它把这些特定序列集合在一起，形成一个蛋白质-DNA的复合体，这个复合体使这些 DNA 位点衔接起来，形成联合复合体（synaptic complex）。在这种复合体中，重组酶催化 DNA 分子的断裂和重新连接，造成 DNA 片段的倒位或位移。一个重组酶通常可以完成所有步骤。这两种类型的重组都被精细地控制着，把在 DNA 中引入断裂和非预期的重组 DNA 片段对细胞带来的危害降到最低。另外，我们将了解到，这两种重组类型在原理上还有一些关键的差异。

下面先介绍比较简单的位点特异性重组，然后对转座进行讨论。每个部分都是先介绍原理的基本特点，然后再提供几个具体的例子。

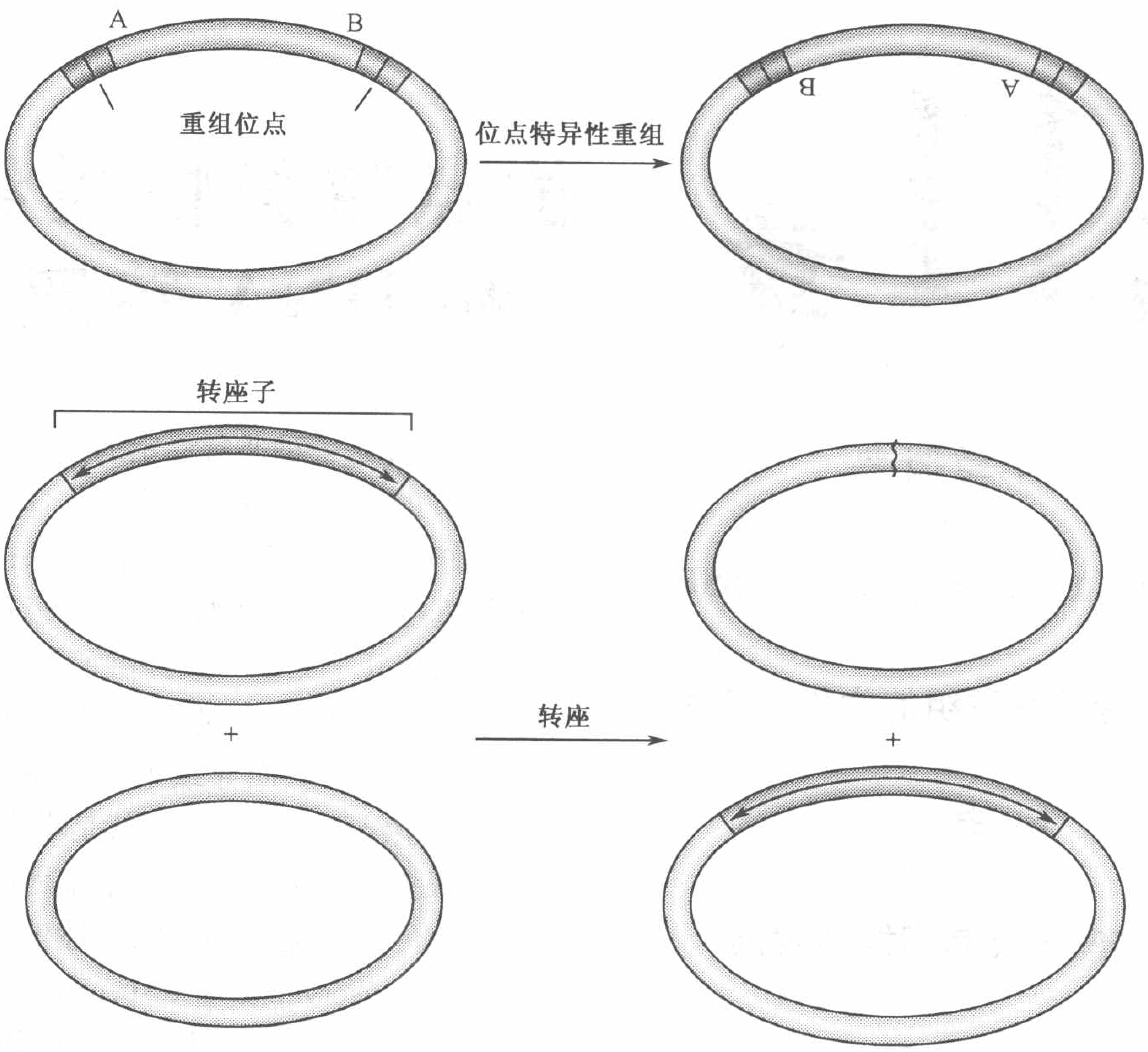


图 11-1 两类遗传重组。上图显示的是位点特异性重组。红色和蓝色重组位点间重组的结果使包含 A 和 B 基因的 DNA 片段发生了颠倒。下图显示的是红色的转座子从灰色的 DNA 上切除并插入到另一个蓝色的 DNA 的不相关位点上。

保守性位点特异性重组（CSSR）

发生在靶 DNA 上特定 DNA 序列的位点特异性重组

很多由一个特定 DNA 片段发生重排的反应都是由保守性位点特异性重组造成的。这些反应的一个关键特性是发生移动的 DNA 片段都带有短而特殊的序列因子，叫做重组位点(recombination site)，在这里可发生 DNA 交换。这种重组类型的一个例子是 λ 噬菌体基因组整合到细菌染色体中（图 11-2 及第 22 章）。

λ 噬菌体整合时，重组总是精确地发生在位于两个重组位点之间一段相同的序列上，这两个重组位点一个在噬菌体 DNA 上，另一个在细菌 DNA 上。重组位点带有两类序列成分：重组酶特异结合的序列和发生 DNA 断裂及重新连接的序列。重组位点往往很短，大约只有 20bp，也可能会更长一些，并带有其他序列基序或蛋白质结合位点。关于更复杂的重组位点的例子，我们将在介绍具体的重组反应时进行说明。

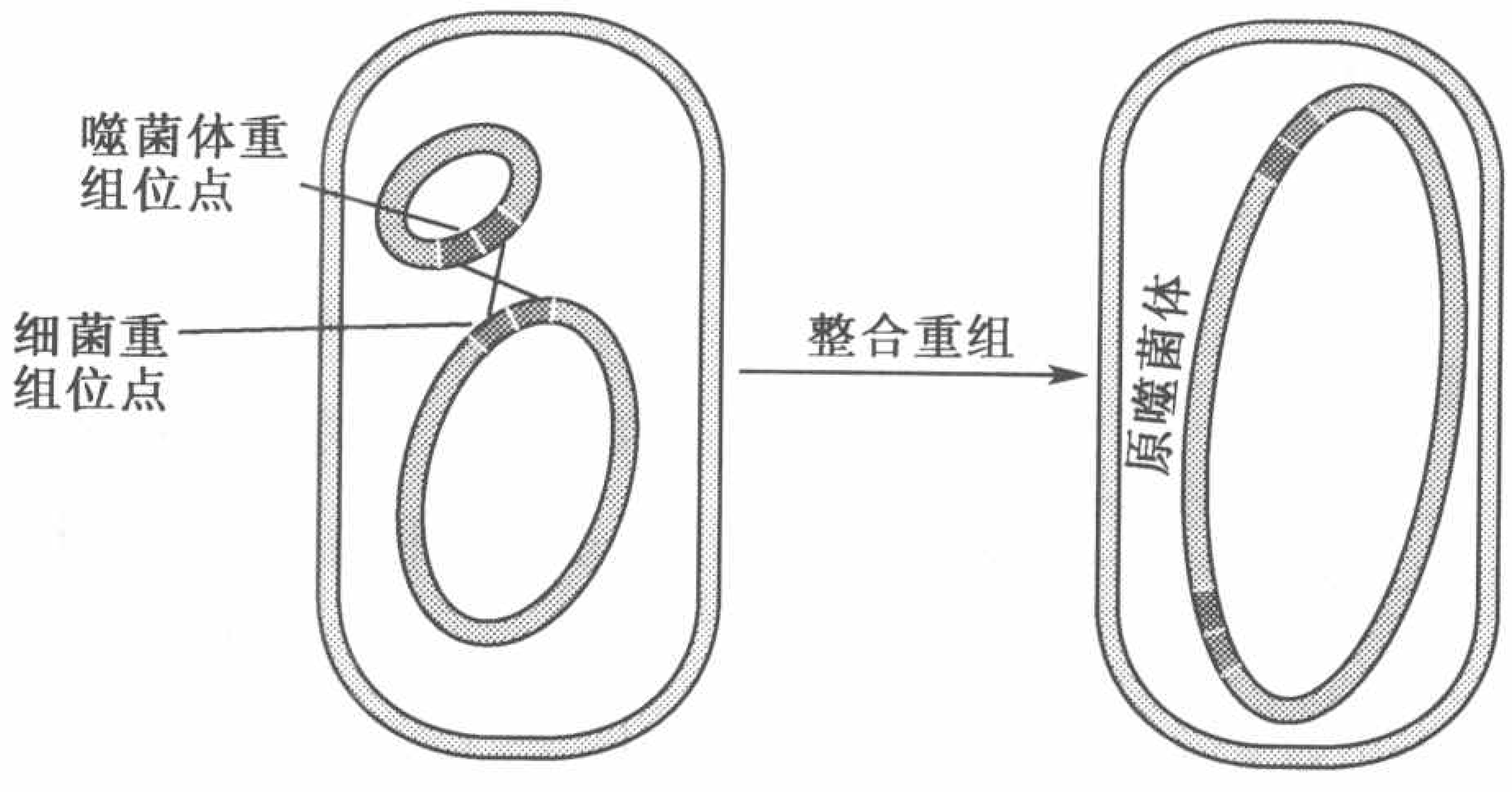


图 11-2 λ噬菌体基因组整合到宿主细胞染色体。在两个 DNA 分子之间 DNA 的交换常特异性地发生在重组位点之间。在图中未按比例显示λ噬菌体和细胞染色体的相对长度。

CSSR 能够产生 3 种不同类型的 DNA 重排（图 11-3）：① 一个 DNA 片段在特定位点的插入（如λ噬菌体 DNA 插入细菌染色体）；② DNA 片段的缺失；③ DNA 片段的倒位。重组结果是造成 DNA 插入、丢失还是倒位，是由 DNA 分子或参与重组的其他分子上的重组识别位点的结构决定的。

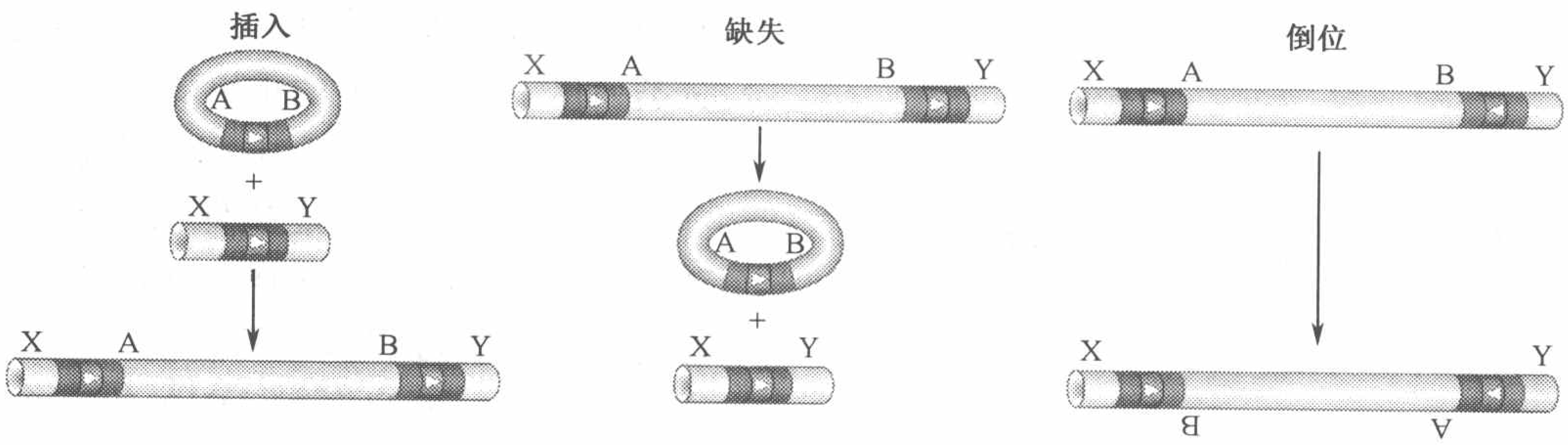


图 11-3 CSSR 重组的 3 种类型。在每个图例中，红色的 DNA 片段通过重组发生了移动或重排。A、B、X 和 Y 代表位于不同 DNA 区段的基因，深红色和蓝色区块代表重组酶识别序列，黑色箭头代表交换区，这些序列一起构成重组位点。

要理解重组位点的结构如何决定 DNA 重排的类型，我们就必须对重组位点的序列成分仔细地研究分析（图 11-4）。每个重组位点由一对对称排列的重组酶识别序列（recombinase recognition sequences）构成，识别序列中间所包围的是一个短的不对称序列，称为交换区（crossover region），DNA 的断裂和重新连接就在这里发生。

由于交换区是不对称的，所以每个重组位点都带有特定的极性。单个 DNA 分子上的两个位点所呈现的方向相关性为两者以反向重复（inverted repeat）或同向重复（direct repeat）的方式排列。在一对反向位点间的重组将使这两个位点间的 DNA 片段发生倒位（图 11-3，右图）；相反，以同样的机制发生在两个同向重复位点间的重组将使这两个位点间的 DNA 片段发生缺失（图 11-3，中图）；最后，插入特异地发生在两个不同分子因 DNA 交换而靠近的情况下（图 11-3，左图）。在对重组酶进行一般性讨论后，将介绍这 3 种重排类型的几个实例。

位点特异性重组酶通过一个蛋白质-DNA 的共价中间体对 DNA 进行切割并重新连接

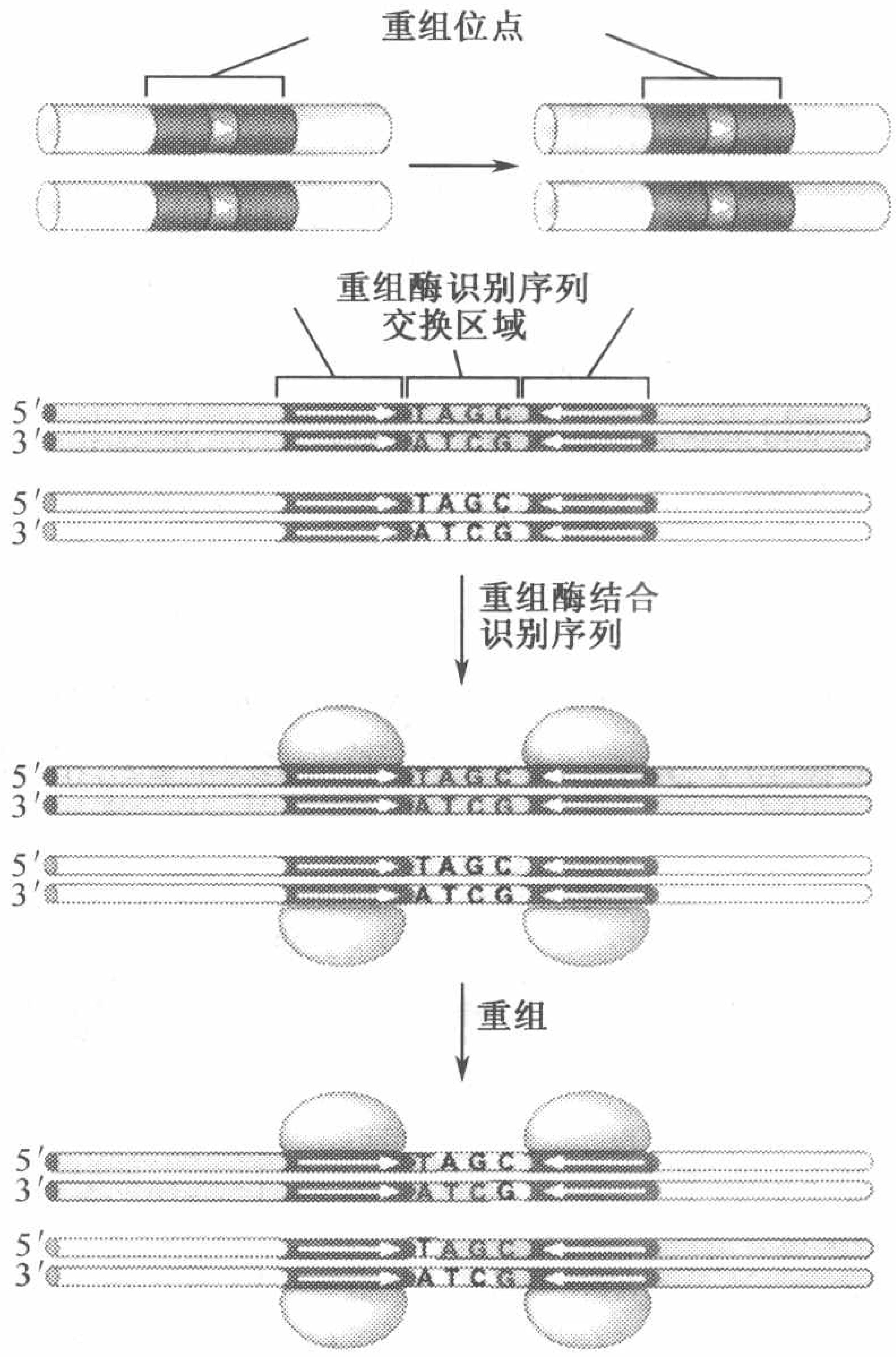


图 11-4 CSSR 过程起作用的结构。一对对称的重组酶识别序列分列在发生重组的交换区两端。重组酶亚基结合到这些重组位点。注意交换区的序列不是回文形式，导致了重组位点内部的不对称。（来源：Craig N. et al. 2002. Mobile DNA Ⅱ，p. 4，Fig. 1. © 2002 ASM Press.）

保守性位点特异性重组酶共有两个家族：丝氨酸重组酶（serine recombinase）和酪氨酸重组酶（tyrosine recombinase）。这两个家族的基本原理都是在切割 DNA 时形成一个蛋白质-DNA 的共价中间体。对丝氨酸蛋白酶而言，该酶活性位点内丝氨酸残基的侧链对重组位点内的特定磷酸二酯键进行攻击（图 11-5），结果在造成 DNA 分子一个单链断裂的同时，在丝氨酸和这个 DNA 断裂位点上的一个磷酸之间形成一个共价连接。同样地，对酪氨酸重组酶来说，则是其活性位点上酪氨酸的侧链攻击 DNA 并与之结合。表 11-1 按家族和生物功能分类列出了许多重要的重组酶。

蛋白质-DNA 共价中间体中蛋白质和 DNA 之间的共价连接保留了断裂的磷酸二酯键的能量。这样，DNA 链可以通过逆转断裂过程被重新连接。在这个逆转过程中，来自断裂 DNA 的一个羟基攻击连接蛋白质和 DNA 的共价键，这一过程共价地缝合了 DNA 缺口，同时释放出游离的（不结合 DNA 的）重组酶（图 11-5）。

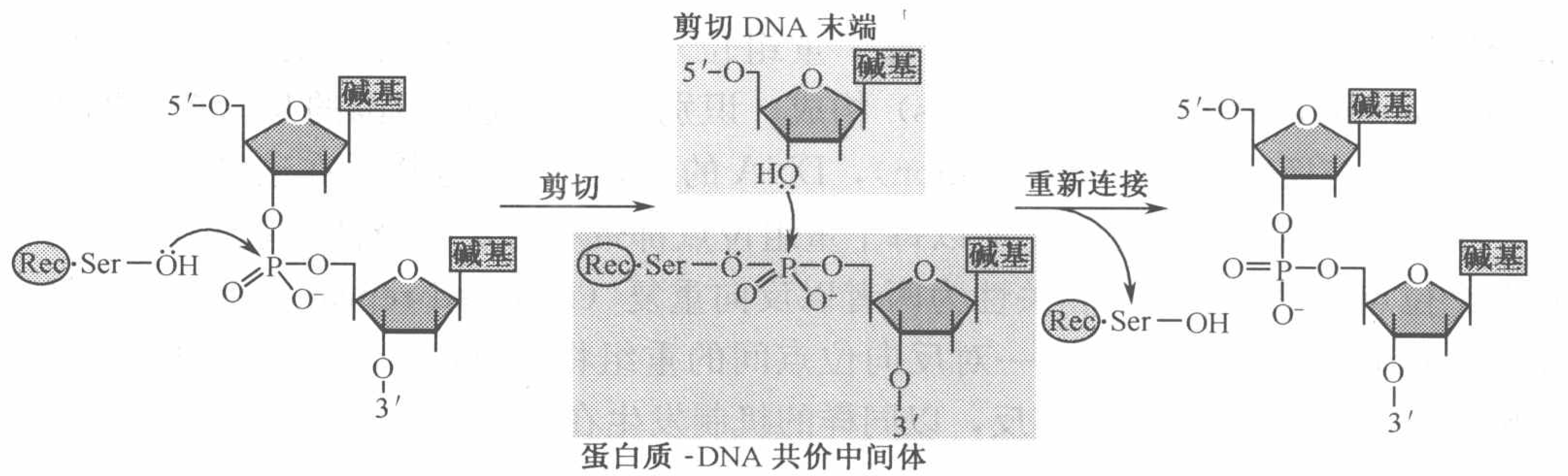


图 11-5 丝氨酸重组酶和酪氨酸重组酶的共价中间体机制。图中所示，丝氨酸活性位点的一个羟基攻击磷酸基团，导致重组位点一条单链的断裂，然后断开的 DNA 上的羟基再攻击蛋白质-DNA 中间的共价键，从而逆转了断裂反应，缝合 DNA，并释放蛋白质。重组酶的标记为 Rec，以蓝色表示。

正是这种机制特性提供了保守性位点特异性重组（CSSR）名称中“保守”的意义：被称为“保守”是因为反应中被断开的每个 DNA 键都被重组酶重新连接起来，而 DNA 在被重组酶断开并重新连接的过程中不需要其他的能量，如 ATP 水解所释放的能量。这种通过共价中间体断裂 DNA 的机制并不局限于重组酶。DNA 拓扑异构酶（第 6 章）和 Spo11 也是这种机制，后者运用这种机制在减数分裂期间诱导 DNA 双链断裂引发同源重组（第 10 章）。

表 11-1　重组酶的家族和功能

重组酶	功　能
丝氨酸重组酶家族	
沙门氏菌 Hin 转化酶	通过识别 *hix* 位点倒位一个染色体区域造成一个基因启动子翻转，表达两种截然不同的表面抗原
Tn3 转座子和 $\gamma\delta$ 解离酶	启动一个 DNA 的删除反应来解除复制性转座造成的 DNA 融合，重组位点称作 *res* 位点
酪氨酸重组酶家族	
λ 噬菌体整合酶	启动 λ 噬菌体基因组整合 λ 或离开大肠杆菌染色体特定序列上的整合及切除反应，重组位点称作 *att* 位点
P1 噬菌体 Cre	通过噬菌体 DNA 上的识别位点（*lox* 位点），在侵染过程中启动噬菌体 DNA 的环化
大肠杆菌 XerC 和 XerD	启动几个 DNA 删除反应，将二聚体的环状 DNA 分子变为单体，同时识别质粒上的（*cer*）位点和细菌染色体的（*dif*）位点
酵母 FLP	将酵母的 2μ 质粒上一个区域倒位以产生一个叫作滚环复制的 DNA 扩增反应，重组位点称作 *frt* 位点

丝氨酸重组酶诱导 DNA 双链断裂及链交换，促成重组

CSSR 总是发生在两个重组位点之间。如上所示，这些位点可以在同一个 DNA 分子上（如倒位和缺失），也可以在两个不同的分子上（如插入）。每个重组位点都是由双链 DNA 构成的。因此，在重组过程中，这 4 条 DNA 单链（每个双螺旋两条）必须先断开再与另一个分子上的单链重新连接，以产生重排的 DNA。

丝氨酸重组酶在进行链交换前将 4 条单链全部切断（图 11-6）。重组酶蛋白中一个分子催化其中一个断裂反应，所以至少需要由 4 个亚基组成（即四聚体形态）重组酶。

亲本 DNA 双链分子的断裂产生 4 个双链 DNA 片段（根据它们所结合的蛋白质的不同分别标记为 R1、R2、R3 和 R4，如图 11-6）。要使重组发生，上边 DNA 分子的 R2 片段必须结合到下边 DNA 分子的 R3 片段上。同样，上边 DNA 分子的 R1 片段必须结合到下边 DNA 分子的 R4 片段上。一旦这一 DNA 交换发生，每个断裂 DNA 链的 3′-OH 端就会攻击新配对片段上的重组酶-DNA 间的共价键。如上所述，这个反应释放了重组酶，而且共价地缝合了 DNA 链间的连接，产生了重排的 DNA 产物。

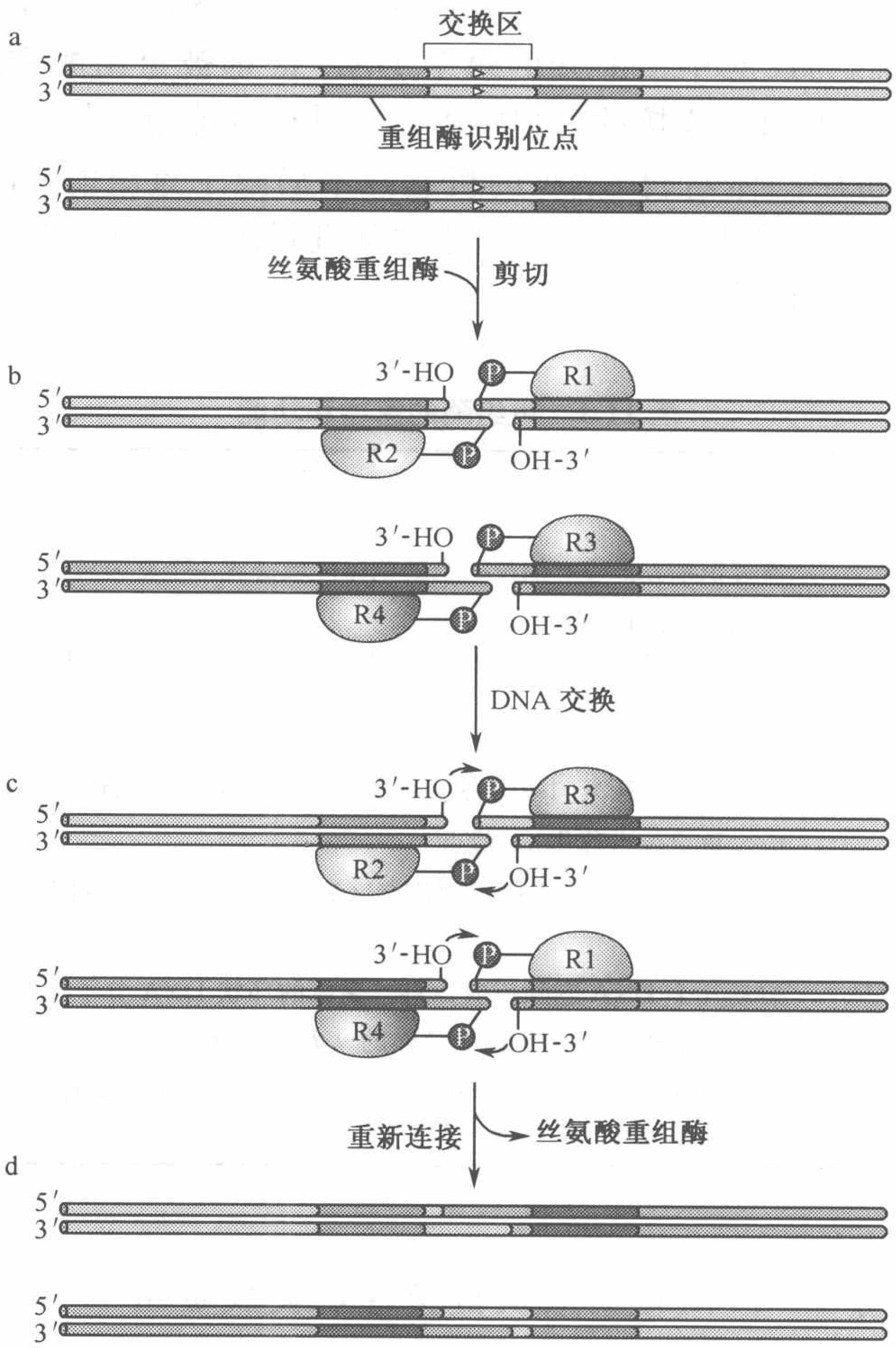


图 11-6 丝氨酸重组酶催化的重组反应。 4 条 DNA 单链各自被重组酶蛋白一个亚基断裂于交换区，4 个亚基被标记为 R1、R2、R3 和 R4。一个双螺旋中的两条单链的断裂位点相差两个碱基，这两个碱基区域在重组产物中形成一个杂交二倍体。图中所示的重组位点与图 11-4 中的类似。

丝氨酸重组酶-DNA 复合体的结构揭示了它通过亚基旋转以完成链交换

重组过程中丝氨酸重组酶-DNA 复合体的结构为 DNA 链如何自然谐调地发生交换提供了一个快照。这个复合体包括丝氨酸重组酶的 4 个亚基和两个断裂的双链 DNA 分子。每个重组酶亚基中的丝氨酸活性位点与每个重组 DNA 位点的 5′磷酸基团之间形成的共价连接清晰可见。每个共价连接的末端依次产生一个带有游离的 3′羟基的 DNA 末端以参与链交换。

这一复合体的最显著特征是在“上面”和“下面”的重组酶二聚体之间具有一个大而平的分界面（图 11-7）。这种结构大多是疏水且光滑的，不太会阻滞复合体上下组分围绕彼此进行旋转。而正负电荷配对的一些区域则使二聚体保持初始状态前趋于

旋转 180°。因此，对于此复合体的分析强有力地证实了重组的机制：（1）DNA 断裂形成共价的酶-DNA 中间体（2）在蛋白-DNA 复合体中二聚体旋转了 180°（3）二聚体旋转后攻击解离酶-DNA 复合体中的 DNA3′羟基末端，使链重新连接形成新的重组构型。

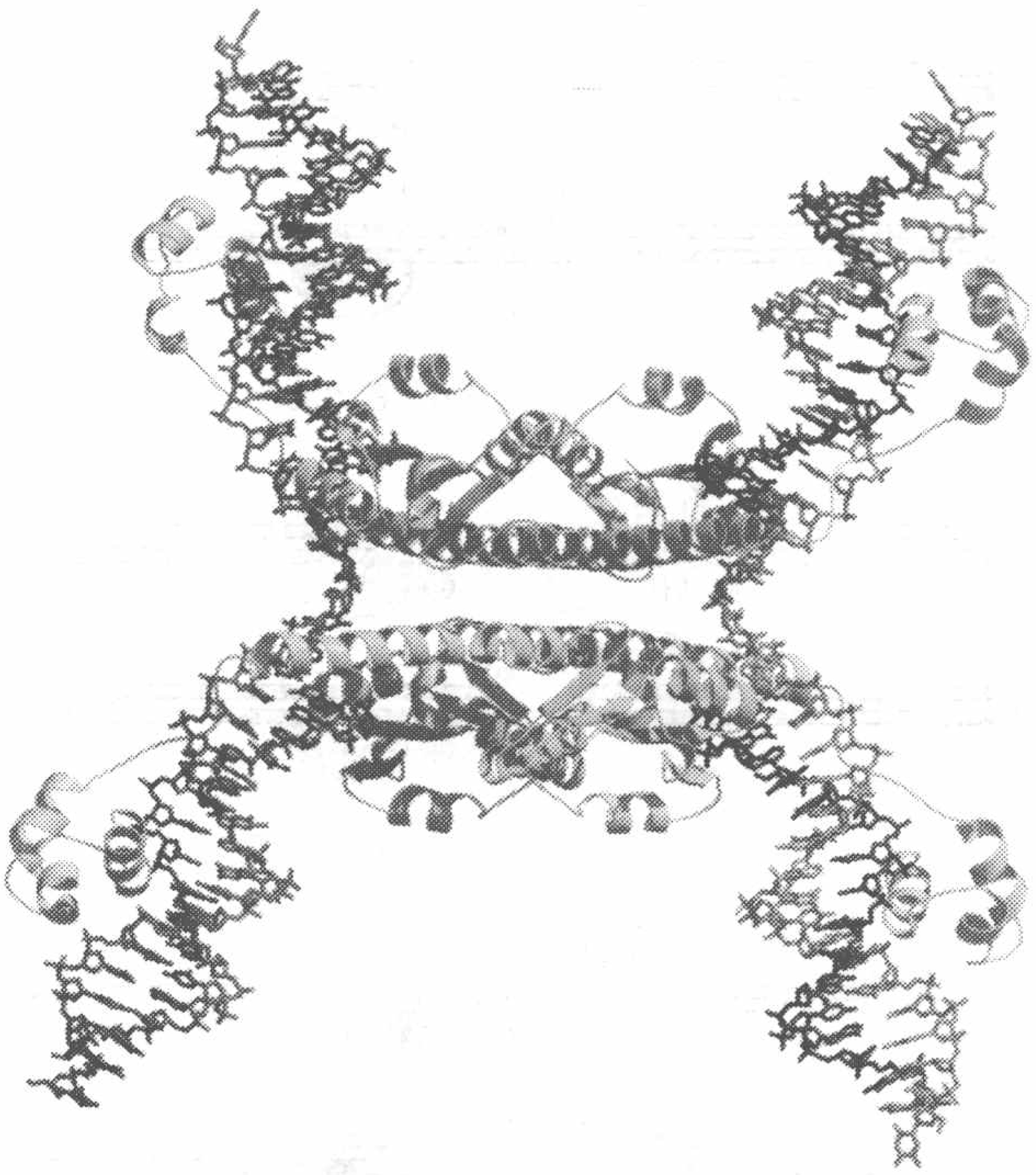


图 11-7 丝氨酸重组酶的结构。这个结构显示的大而平的四聚体界面是旋转的位点。尽管在这里只显示了两条链被切断，对于一个完整的重组反应，4 个 DNA 链必须全部断裂（如图 11-6）。（Li W. et al. 2005. Science 309：1210. PDB Code：1ZR4.）图形通过 MolScript，BobScript 和 Raster 3D 制作。

酪氨酸重组酶每次断裂和结合一对 DNA 单链

和丝氨酸重组酶不同，酪氨酸重组酶先完成两条 DNA 单链的断裂和重新连接，再对剩下的两条单链进行断裂和重新连接（图 11-8）。考虑到是两个 DNA 分子与它们的重组位点配对，因此同样需要 4 个分子的重组酶，每个分子断裂一条 DNA 单链。重组开始时，结合到左侧重组酶结合位点的亚基（在图 11-8a 中标记为 R1 和 R3）分别断裂各自结合的 DNA 分子上的单链，断裂发生在交换区的第一个核苷酸处。然后，上边灰色 DNA 分子的右上单链和下边红色 DNA 分子的右上单链交换伙伴。在重组构型中，这两条 DNA 单链被连接到一起。这个“第一单链”交换反应所产生的一个带分岔的 DNA 中间体称为 Holliday 联结体（见第 10 章，图 11-8b）。

一旦第一单链交换完成，另外两个重组酶亚基（标记为 R2 和 R4）就会切割两个 DNA 分子的下单链（图 11-8c）。它们再交换并经过与断裂反应逆转的过程连接起来。

这个"第二单链"的交换反应打开了 Holliday 联结体（见第 10 章），得到重排的 DNA 产物。下一节我们将讨论这些化学反应是如何在重组酶蛋白质-DNA 复合体的作用下发生的。

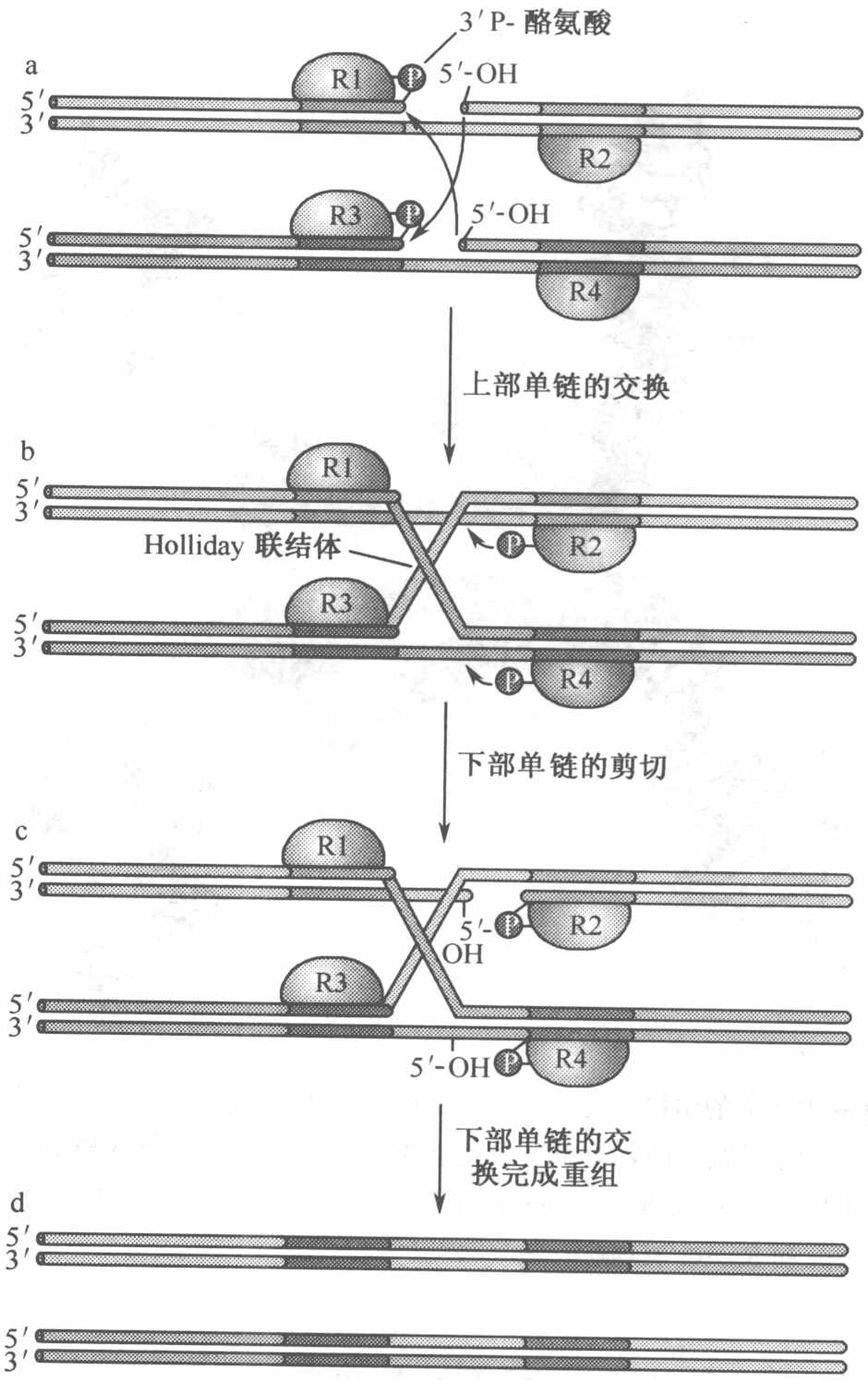


图 11-8 酪氨酸重组酶催化的重组反应。第一步，由 R1 和 R3 亚基断裂 DNA（a）。如图所示，蛋白质与断裂的 DNA 通过 3′端磷酸-酪氨酸键连接。断裂 DNA 5′端的两个羟基各自攻击另一个 DNA 分子上蛋白质-DNA 的连接键后发生第一对单链的交换（b）。第二条单链的交换机制与上面相同，通过 R2、R4 亚基完成（c 和 d）。（来源：Craig N. et al. 2002. Mobile DNA Ⅱ，color plate 1，chapter 2. © 2002 ASM Press.）

酪氨酸重组酶与 DNA 结合的结构揭示了 DNA 交换的机制

酪氨酸重组酶的位点特异性重组机制已经了解得很清楚了。这一蛋白质家族的一些成员的结构已经得到了很好的了解，并揭示出其在重组过程中的作用。一个完美的例子就是 Cre 重组酶结合到重组 DNA 两种不同构型上时的结构。下面介绍从这些结构得出的详细机制。Cre 是 P1 噬菌体编码的一种酶，其功能是在侵染过程中环化线性的噬菌体基因组。Cre 作用于 DNA 上的重组位点叫 *lox*，Cre-*lox* 是一个由酪氨酸重组酶家族

催化重组的简单例子，整个重组仅需 Cre 蛋白和 *lox* 位点。Cre 也是在遗传工程中被广泛使用的一个工具（框 11-1 位点特异性重组在遗传工程中的应用）。

框 11-1 位点特异性重组在遗传工程中的应用

由于一些位点特异性重组系统非常简单，因而被广泛用作遗传实验的工具。Cre 重组酶和其“近亲”FLP 重组酶，都被用在真核生物基因删除的实验中（见第 22 章的例子）。

我们举一个假设的例子就能清楚地说明这种策略的益处。例如，一个研究人员对某个基因在肺癌发育过程中的作用感兴趣，她希望用小鼠作为模型生物来研究这个过程（第 22 章）。当该基因被破坏（“敲除”）后（见图 22-27），小鼠在胚胎形成早期就都死了。很明显，这个基因在小鼠早期发育中是必需的。那么，该如何研究该基因在成年动物肺癌形成中的作用呢？

这时位点特异性重组往往可为我们提供答案。通过传统的方法在感兴趣的基因两侧引入可被 Cre（或 FLP）识别的重组位点，在没有重组酶的情况下，这些位点对该基因的功能不造成任何影响。因此，可将编码 Cre 蛋白（或 FLP 蛋白）的基因也引入该生物体内，并将其置于一个启动子的严格控制之下（第 17 章）。这样，小鼠先在无重组酶出现的情况下发育，出生之后，“开启”Cre 的表达，重组酶的出现就造成了该基因的丢失。通过这种方法，把 Cre 处理（将感兴趣的基因删除）后的小鼠的情况与该基因保持完整的“正常”小鼠进行比较，就可以进行该基因在肺癌发育过程中的作用。这样，利用 Cre 形成的重组使得我们可以在不同的发育时期研究基因的潜在功能。

Cre-*lox* 结构表明，重组需要 4 个 Cre 亚基，每个亚基结合到底物 DNA 上的一个结合位点（图 11-9）。DNA 的构象通常是一个四路交叉的十字路口结构（第 10 章关于 Holliday联结体的讨论），每个“臂”上都结合了一个 Cre 亚基。尽管乍看起来这个结构的四边是对称的，但实际上不是这样。Cre 有两种不同的构象，构象 1 的一对亚基用绿色标记，构象 2 的另一对亚基用粉色标记（图 11-9b）。只有一种构象（绿色标记的亚基）能够切断和重新连接 DNA。因此，每次只有一对亚基处于活动状态。随着反应的进行，激活构象在两对亚基间来回交换，而这种交换对于控制重组过程的进行和保证“每次作用一条单链”的交换机制是至关重要的。

位点特异性重组的生物学作用

细胞和病毒利用保守性位点特异性重组来实现多种多样的生物学功能，某些功能将在下面进行讨论。很多噬菌体的感染就是使用这种重组机制将自己的 DNA 插入到宿主染色体中。还有些时候，位点特异性重组被用来改变基因的表达。例如，一个 DNA 片段的倒换可以决定两个基因中哪一个进行表达。位点特异性重组也被广泛地用于帮助环状 DNA 分子在复制、同源重组和细胞分裂过程中保持其结构完整性。

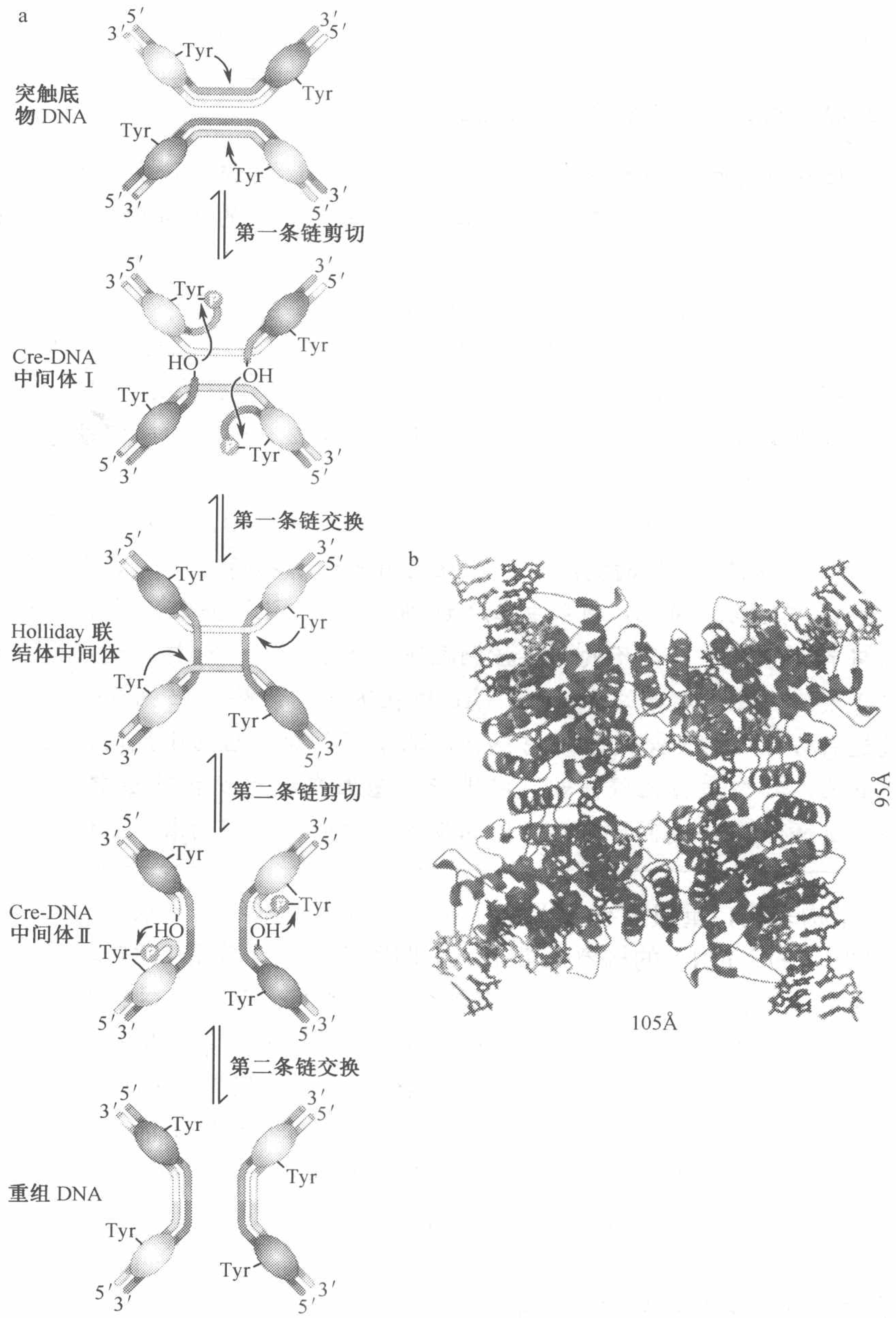


图 11-9 Cre 重组酶的位点特异性重组机制。 a. 左图是一系列 Cre-DNA 中间体的结构，它反映了“每次作用一条单链”的交换机制。在每个分图中，只有两个绿色的亚基是活性构象。可以看出，当第一条单链断裂后亚基的颜色发生了交换，第二对 Cre 亚基变为重组的活性构象。（来源：Guo F. et al. 1997. Nature 389：41. @ Macmillan.）b. 右图所示为有 Cre 结合时 Holliday 联结中间体的晶体结构（对应于图 a 的第三张图）。注意绿色的两个亚基的构形与紫色的不同，因此复合体不是一个对称的四边形。留心观察可以发现，结构中相邻的两对 DNA 臂更加接近一些。（Gopaul D. N. et al. 1998. EMBO J. 17：4175.）图形通过 BobScript、MolScript 和 Raster 3D 软件制作。

位点特异性重组系统之间的比较揭示出一些普遍的原理。所有反应的关键是重组酶与DNA的结合，以及使两个重组位点聚合在一起。对于一些重组反应来说，这个结合过程很简单，仅需要重组酶和它的DNA识别序列，就像刚描述的Cre的情况一样。与此相反，另外一些重组反应则需要辅助蛋白。这些辅助蛋白包括一些所谓的构筑蛋白（architectural protein），它们结合到特定DNA序列上并使其弯曲，使之成为有利于重组过程进行的特异形状。构筑蛋白还可以控制重组反应发生的方向，例如，在阻止DNA切除这个反向反应的同时确保这个DNA片段的整合。显而易见，这种调节作用对于产生合乎逻辑的生物学结果是至关重要的。最后我们会了解到，通过其他蛋白质对重组酶的调节可以控制某种DNA重组发生的时间，并使其与其他细胞活动协调一致。

λ整合酶催化病毒基因组在宿主细胞染色体上的整合和切除

当λ噬菌体侵染宿主细菌时，一系列的调节活动导致其或者形成静止的溶原状态（lysogenic state），或者进行噬菌体的增殖，即进入裂解性生长（lytic growth）过程（第16和第22章）。溶原体的形成需要噬菌体DNA整合到宿主染色体上。同样地，要使噬菌体脱离溶原状态，开始复制并产生新的噬菌体颗粒，它就必须将自己的DNA从宿主染色体上切下来。对这一整合/切除过程的分析为我们提供了从分子水平上对位点特异性重组进行研究的切入口。

为实现整合，整合酶蛋白（λInt）在两个特定位点之间催化重组的进行，这两个位点叫做*att*位点，或者RH附着位点。*attP*位点在噬菌体上（P代表噬菌体），*attB*位点在细菌染色体上（*B*代表细菌，图11-2）。λInt是一个酪氨酸重组酶，所催化的单链交换机制遵循前面介绍的Cre蛋白的催化过程。然而，与Cre重组不同的是，λ整合需要辅助蛋白的帮助来组装所需的蛋白质-DNA复合体。辅助蛋白控制着反应过程，确保在噬菌体生命周期的正确时间进行DNA的整合和切除。我们先介绍整合的过程，再了解切除是如何被启动。

对λ整合的调节中很重要的一点是*attP*和*attB*位点的高度不对称结构（图11-10）。这两个位点都含有一个中央核心片段（约30bp），这些核心重组位点都包含有两个λInt结合位点和一个发生单链交换的交换区（如前所述）。不同的是*attB*仅由这个中央核心区组成，而*attP*则长很多（240bp），它还带有许多其他蛋白质的结合位点。

*attP*核心区两端的DNA区域被称为“臂”，它们带有若干个蛋白质结合位点，包括λInt结合的其他位点（在图11-10中标记为P_1、P_2和P'）。λInt是一个不寻常的蛋白质，因为它有两个结构域参与了序列特异性的DNA结合：一个结构域结合到臂上的重组酶识别位点；另一个结合到核心识别位点。另外，*attP*臂上的位点还被几个构筑蛋白所结合，这些蛋白质的结合控制了重组进行的方向和效率。

整合过程要求*attB*、*attP*、λInt，以及一个叫做整合宿主因子（integration host factor，IHF）的构筑蛋白的参与。IHF是一个序列依赖性的DNA结合蛋白，它能使DNA产生大角度（大于160°）的弯曲(图11-11)。*attP*的臂上携带了3个IHF位点（在图11-10上被标记为H_1、H_2和H'）。IHF的作用是将DNA臂上的λInt位点（λInt与之结合较强）与中央核心区上的位点（λInt与之结合较弱）聚合在一起。因此，由

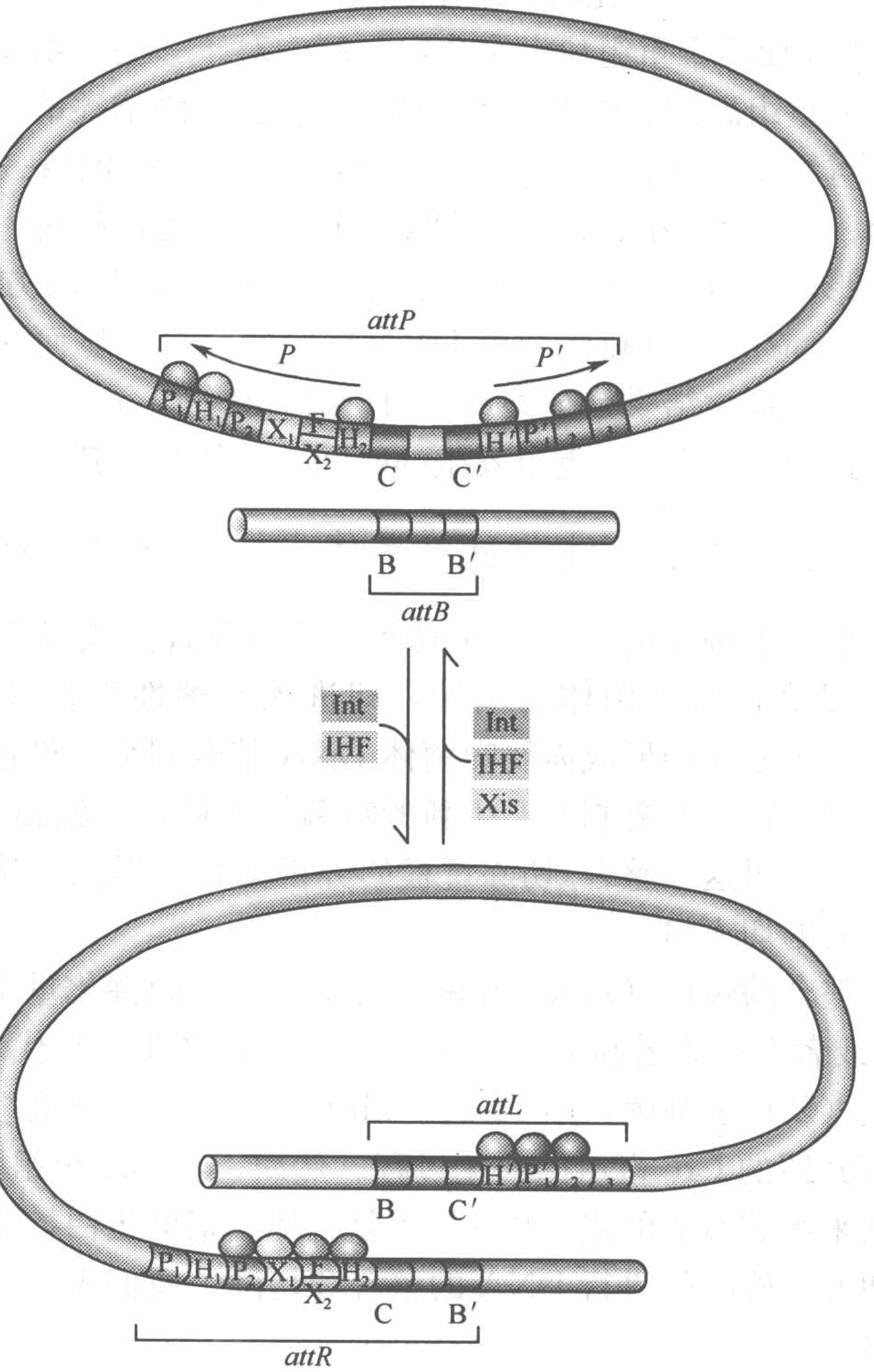


图 11-10 参与 λ 噬菌体整合和切除的重组位点显示出的重要序列因子。 C、C′、B 和 B′是核心的 λInt 结合位点，其他的蛋白质结合位点位于 *attP* 上及 C 和 C′位点两端。这些区域叫做“臂”，左边的叫 P 臂，右边的叫 P′臂。标记为 P_1、P_2 和 P'_1 的紫色小区块是臂上的 λInt 结合位点。H 是 IHF 的结合位点，X 是 Xis 的结合位点，F 是 Fis 的结合位点，其他的构筑蛋白在这里不再作介绍。灰色区域是交换区。为了表示清楚，结合在核心位点上的 λInt 没有画出来。注意在结合或切除的重组过程中，并不是所有的蛋白质结合位点都有蛋白质结合。重组完成后，P 臂成为 *attL* 的一部分，而 P′臂变为 *attR* 的一部分。

IHF 介导使 DNA 产生大角度弯曲，使 λInt 找到结合能力较弱的核心识别位点，从而催化重组的发生。

当重组完成后，环形的噬菌体基因组稳定地整合到宿主染色体上。结果，在噬菌体和宿主 DNA 的结合处产生了两个新的杂交位点。这些位点被称为 *attL*（左侧）和 *attR*（右侧）（图 11-10）。这些位点都含有核心区域，但是臂区却被分开了（图 11-10 中 P 和 P′区域的位置）。所以，在新的重排结构中，两个核心区域都没有能力运用产生整合复合体的机制来组装有活性的 λInt 重组酶复合体，因为对组装很关键的 DNA 位点根本不

在正确的位置上。

λ细菌噬菌体的切除需要一个新的折叠DNA的蛋白质

λ噬菌体是如何被切除的呢？一个由噬菌体编码的构筑蛋白对切除重组至关重要，这个蛋白质叫做Xis（即切除），它结合到特定的DNA序列上并造成DNA的弯曲。这样看来，其功能与IHF类似。Xis识别*attR*一个臂上的两个主要序列单元（也存在于*attP*臂上，如DNA图11-10中标记为X_1和X_2）。和这些序列结合后，DNA产生一个大的弯曲（大于140°），通过聚合Xis、λInt和IHF一起构成在*attR*上有活性的蛋白质-DNA复合体来促进切除过程。这个复合体将富有成效地与聚集在*attL*上的蛋白质发生相互作用，然后发生重组。

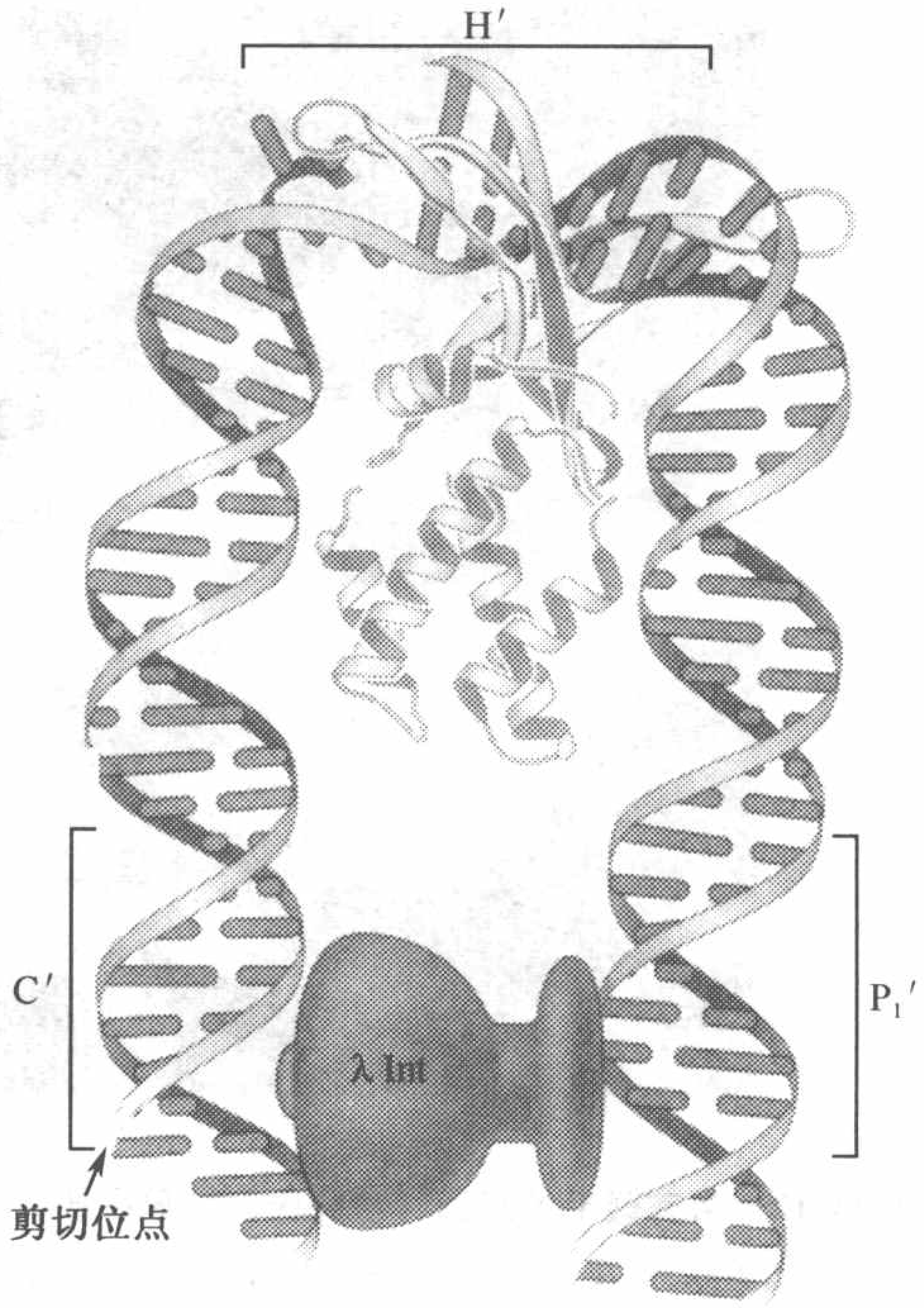


图11-11 使DNA弯曲以便将DNA结合位点聚合在一起的IHF作用模型。 位于*attP*的P′臂上的λInt和IHF结合位点如图所示。IHF结合到H′位点后将DNA弯曲，使一个分子的λInt同时与P_1'和C′位点结合。DNA上H′位点中的断裂表现为在DNA上出现一个缺口，并被用于IHF-DNA复合体的结构分析。（来源：Rice P. et al. 1996. Cell 87：1295-1303. Fig. 8. © Elsevier.）

除了促进切除过程（*attL*和*attR*之间的重组）以外，Xis结合到DNA上也抑制了整合过程（*attP*和*attB*之间的重组）。因为Xis和*attP*结合后的DNA结构与λInt和IHF正常结合到这些位点时的结构完全不同。Xis是一个噬菌体编码蛋白，而且仅仅在噬菌体启动进入裂解性生长的情况下才开始表达。关于Xis表达的详细情况将在第16章中介绍。Xis在切除过程中作为辅助因子起到刺激作用，在整合过程中起到抑制因子的作用，这种双重功能保证了在Xis存在的情况下噬菌体基因组能够从宿主染色体中被切除，并保持游离的状态。

Hin重组酶颠倒一段DNA片段，使得特定基因开始表达

沙门氏菌（*Salmonella*）的Hin重组酶通过将一个细菌染色体片段倒位，来控制两组基因中哪一组被表达。Hin重组是一类重组反应的代表，这些重组在细菌中是相当普遍的，被称为程序性重排。它的功能通常是让一部分细菌能够对周围环境的突然变化获得“预适应”。在Hin倒位的情况中，重组能够帮助细菌躲避宿主免疫系统的攻击，我们即将对此进行说明。

倒位过程所控制的基因能编码两种不同形态的鞭毛蛋白（H1和H2），它们是鞭毛纤丝的蛋白质组分。鞭毛位于细菌表面，因而通常是免疫系统的识别目标(图11-12)。通过用Hin重组实现在这两种形态间的切换，至少使细菌群体中的部分个体不被免疫系统识别出其表面的结构。

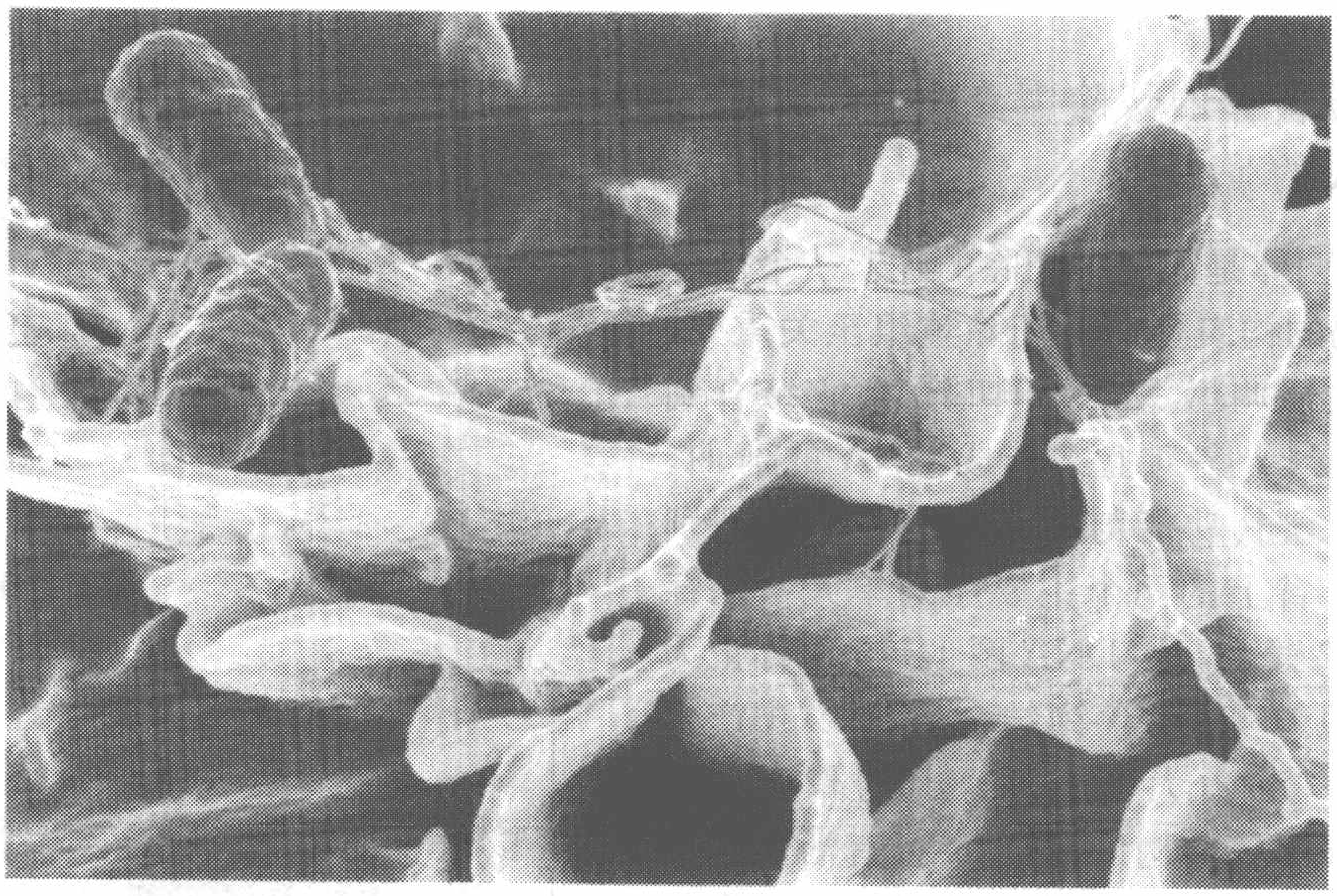


图 11-12 带有鞭毛的细菌（沙门氏菌）显微照片。 这张彩色的增强型电子扫描显微图片显示了鼠伤寒沙门氏菌（红色）正在侵入培养的人体细胞。细菌上发丝状的突起就是鞭毛。（来源：Courtesy of the Rocky Mountain Laboratories，NIAID，NIH.）

被 Hin 倒位的染色体区域大约长 1000bp，其两端是叫做 *hixL*（左侧）和 *hixR*（右侧）的特异性重组位点（图 11-13），这些序列彼此间方向相反。Hin 是一种丝氨酸重组酶，它催化倒位的发生，工作原理在前面介绍这个家族时已经讲过了。在可倒位的片段上带有编码 Hin 的基因，同时作为一个启动子，当可倒位片段处于某种方向时，它将启动位于可倒位片段外、与 *hixR* 位点直接相邻的一些基因的表达。当可倒位片段处于“开”的方向时，这些相邻的基因将得到表达；但当它翻转成“关”的方向时，由于缺乏有功能的启动子，这些基因无法被转录。

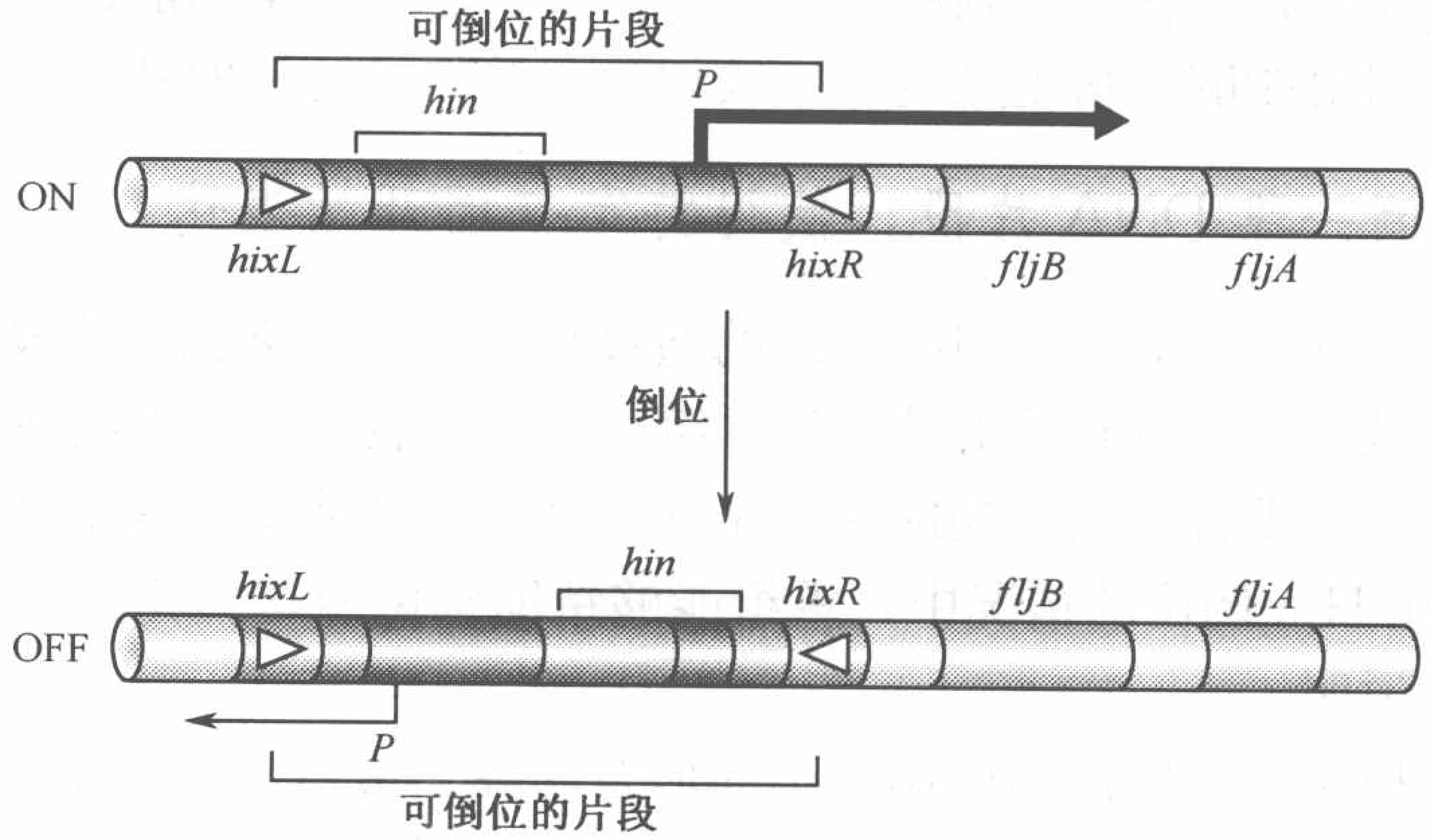


图 11-13 沙门氏菌 Hin 重组酶催化的 DNA 倒位。 Hix 位点间 DNA 片段的倒位翻转了一个启动子（P）的位置，造成了两种不同形式的鞭毛基因表达。

被这个“翻跟头”的启动子所控制的两个基因是 *fljB* 和 *fljA*。*fljB* 编码 H2 鞭毛蛋白，*fljA* 则编码针对 H1 鞭毛蛋白基因的转录抑制因子。H1 鞭毛蛋白基因位于较远的一个位点。这样，当在“开”的方向时，H2 鞭毛蛋白和 H1 的抑制因子被表达，这些细胞的表面毫无例外都是 H2 型鞭毛蛋白；而当处于“关”的方向时，H2 和 H1 的抑制因子都不表达，细胞表面只有 H1 型的鞭毛。

Hin 重组过程需要一个 DNA 增强子

Hin 重组过程除了需要 *hix* 位点外，还需要一段序列。这段短短的序列（约 60bp）是一个增强子，它能将重组率提高 1000 倍左右。像增强子促进转录的作用（第 17 章）一样，这段序列即使在离重组位点很远的情况下也能行使其功能。增强子要起作用还需要细菌的 Fis 蛋白［作为倒位促进因子（factor for inversion stimulation，Fis）被发现而得名］。像 IHF 一样，Fis 也是一个位点特异性的 DNA 弯曲蛋白。另外，它与 Hin 蛋白发生接触，这对重组过程很重要。

增强子-Fis 蛋白的复合体激活了重组的催化步骤。Hin 可以在没有 Fis-增强子复合体存在的情况下与 *hix* 重组位点结合配对，形成突触复合体（图 11-14）。这与 λ 整合中 IHF 的作用不同，后一过程中辅助蛋白 IHF 在重组酶-DNA 复合体的组装中是不可或缺的。Fis 激活 Hin 重组，需要 3 个 DNA 位点（*hixL*、*hixR* 和增强子）聚集在一起，而负 DNA 超螺旋（第 6 章）极大地促进这三联复合体的形成，它能够稳定远距离 DNA 位点的聚合。另一个细菌构筑蛋白 Hu 也能促进这个倒位复合体的组装。Hu 蛋白与 IHF 的结构很相似，但与 IHF 不同的是它与 DNA 的结合是非序列特异性的。

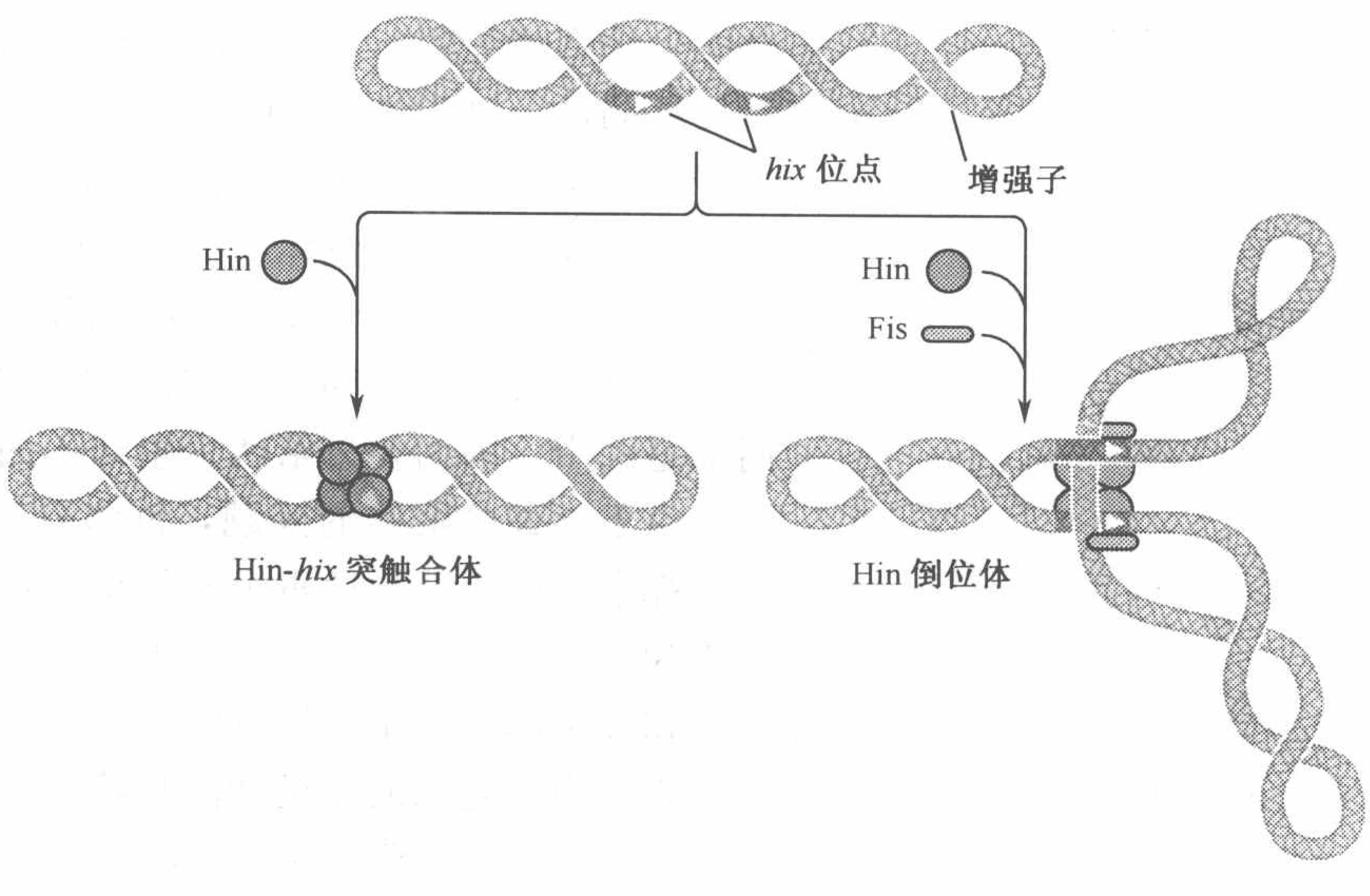


图 11-14 Hin 催化过程中重组复合体的形成。Hin 蛋白能单独识别并与两个 *hix* 位点配对。当 Fis 蛋白存在时，就可以形成三元复合体，这个复合体叫做倒位体（invertasome），是促进重组的最活跃的复合体。（来源：Craig N. et al. 2002. Mobile DNA Ⅱ，p. 246，Fig. 9. © 2002 Copyright ASM Press.）

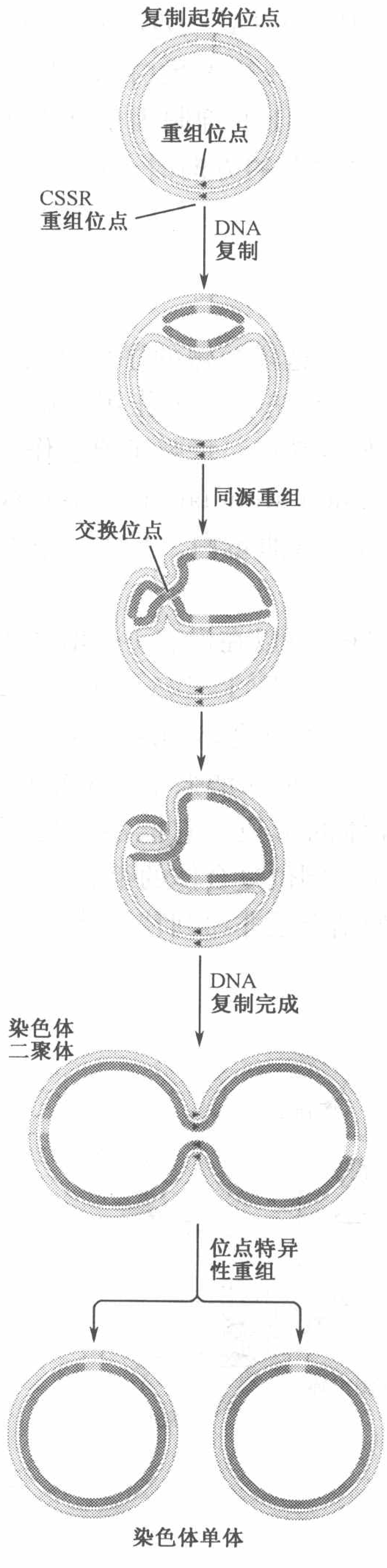


图 11-15 环形 DNA 能形成多聚体。DNA 复制期间，两个子代 DNA 分子间的同源重组能产生一个染色体（或质粒）的二聚体。XerCD 重组酶催化的位点特异性重组将其变为细胞正常分裂所需的单倍体 DNA 分子。

是什么生物学基本原理控制着由 Fis-增强子复合体催化的 Hin 倒位呢？那就是确保重组仅仅发生在同一个 DNA 分子的两个 *hix* 位点之间。这种选择性要确保的是避免可倒位片段在 DNA 分子间重排过程中频繁跳跃而造成破坏细菌染色体的完整性。

与 λ 噬菌体的整合和切除不同，Hin 所催化的倒位没有严格的调控，反而是随机发生的，所以在一群细胞中总有一些细胞带有二个方向的可倒位片段，而另一些带相反方向的可倒位片段。

重组酶将多聚环状 DNA 分子转换成单体

位点特异性重组对细胞中环状 DNA 分子的维护起到关键作用。如同原核和真核细胞中许多质粒一样，大多数细菌的染色体也是环状的，一些病毒基因组也是环状的。环状 DNA 分子固有的一个问题是，在同源重组过程中它们有时会形成二聚体甚至多聚体。而位点特异性重组可被用来将这些 DNA 多聚体恢复到单体形态。

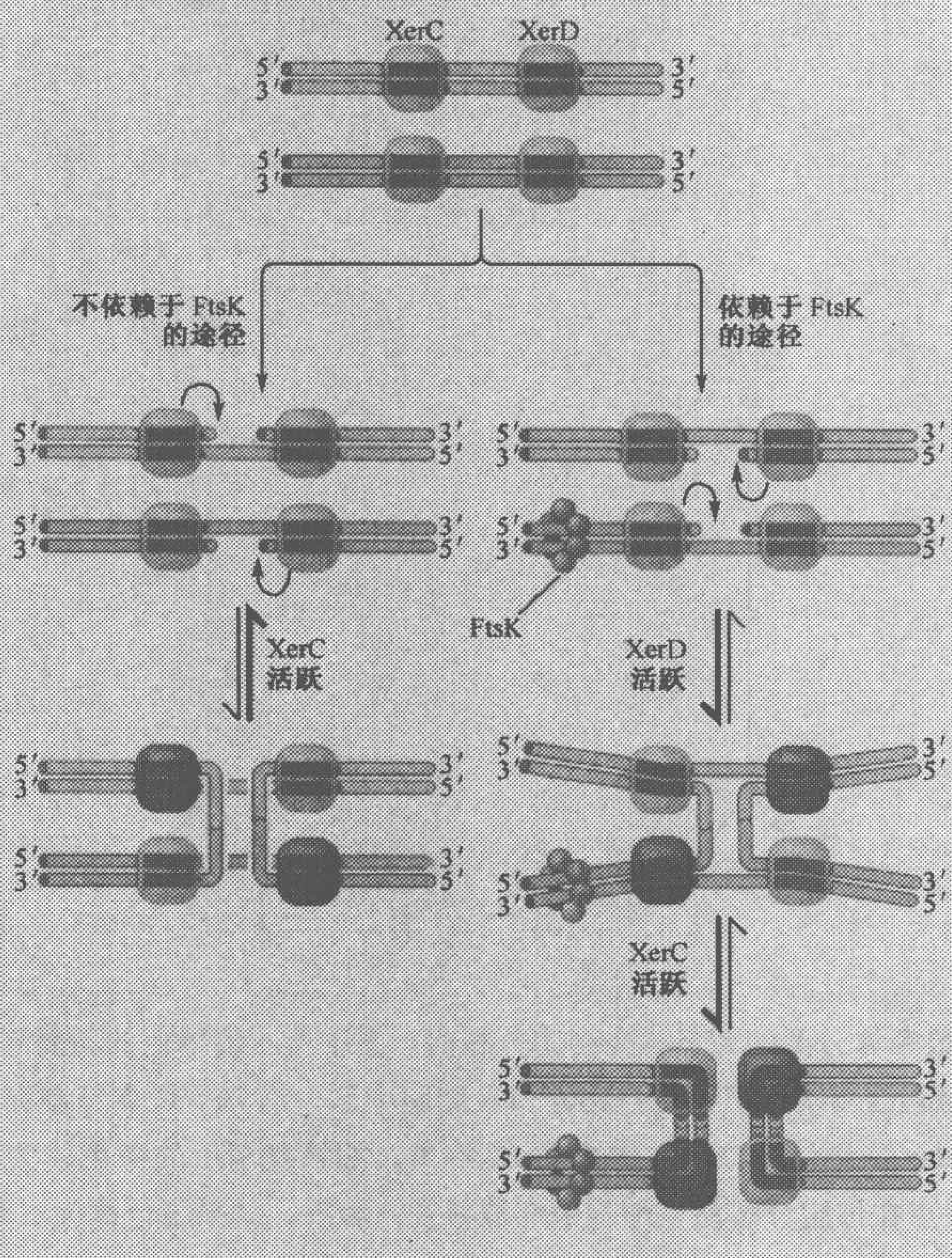
当两个相同的环状分子间发生 DNA 交换时会发生什么情况呢？在一个细菌染色体两个拷贝间发生的这一过程如图 11-15 所示（见第 10 章，关于同源重组的讨论）。一个单独的同源重组会产生一个大的环状染色体，所有基因都有两份拷贝，即一个二聚的染色体。在细胞分裂时，这种二聚体会造成很大麻烦，只有一个 DNA 分子，而不是两个。因为要分给两个姐妹子细胞。

由于这种多聚体的问题，许多环形 DNA 分子上都带有位点特异性重组酶的识别位点。这些酶称为解离酶（resolvase），因为它们能将二聚体（和多聚体）“解离”为单体。显然，对这些蛋白质来说，它们的关键功能是特异性地催化解离（一种 DNA 的删除反应）而不是相反的过程（将单体变为多聚体），否则只能使多聚体问题变得更糟！我们将会看到，一种特殊机制在增强重组反应中起到了方向选择的作用。（框 11-2）

框 11-2 Xer 重组酶催化细菌染色体和很多细菌质粒的单体化

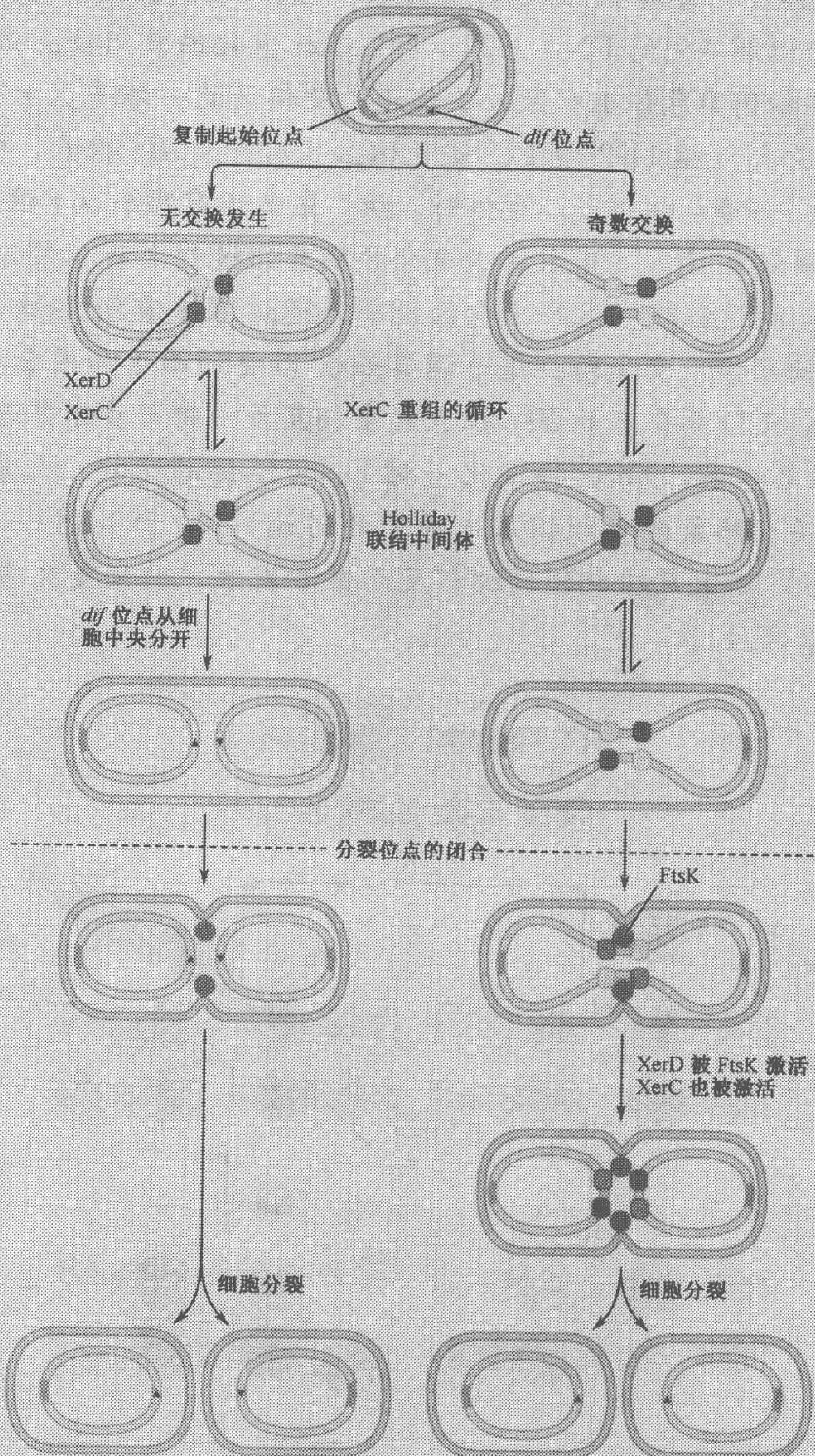
Xer 是酪氨酸重组酶家族的一个成员，它的重组机制与前面讲过的 Cre 蛋白很相似。Xer 是一个异源四聚体，包含两个 XerC 蛋白亚基和两个 XerD 蛋白亚基。XerC 和 XerD 都是酪氨酸重组酶，但识别不同的 DNA 序列。所以 Xer 催化的重组位点必须要包含这两种识别序列。该位点在细菌染色体上叫做 *dif* 位点。交换区的一端是 XerC 的识别序列，另一端是 XerD 的识别序列（框 11-2 图 1）。染色体上只有一个 *dif* 位点，它位于 DNA 复制终止区内（第 8 章）。当染色体形成二聚体时，该二聚体便有两个 *dif* 位点（图 11-15）。

那么细胞如何确保在 *dif* 位点上由 Xer 介导的重组过程会将二聚体染色体转换为单体，而不是促成相反的过程呢？这个定向的调节是通过 Xer 重组酶和一个叫 FtsK 的细胞分裂蛋白之间的相互作用实现的。这个调节如框 11-2 的图 1 和图 2 所示，其过程是：当 FtsK 蛋白不与 XerCD 复合体的 *dif* 位点发生相互作用时，重组复合体呈仅具有两个 XerC 亚基活性的构象结果，XerC 将催化一对 DNA 单链的交换，形成Holliday联结体中间体（见前面的关于酪氨酸重组酶基本原理的讨论）。由于 XerD 一直没有激活，重组无法完成，与 XerC 断裂反应相逆的过程反而常常发生，结果是又重新产生了最初的 DNA 结构（框 11-2 图 1）。



框 11-2 图 1　Xer 介导 *dif* 位点重组的途径。没有 FtsK 存在时（不依赖于 FtsK 途径，左图），仅 XerC 激活并推动单链交换形成 Holliday 联结中间体，这时（由于 XerD 没有激活）重组无法完成，XerC 的反应也常常发生逆转。当有 FtsK 存在时（依赖 FtsK 途径，右图），XerD 被激活并催化形成 Holliday 联结中间体，XerC 催化第二对单链的交换后完成重组，生成染色体单体。（来源：Aussel L. et al. 2002. Cell 108：195-205，Fig. 6 @ Elsevier.）

相反，当 FtsK 蛋白存在并与 XerCD 复合体发生相互作用，复合体的构象发生改变，XerD 蛋白被激活。这时，XerD 催化第一对单链的重组，产生 Holliday 联结体；这一反应完成后，XerC 催化第二对单链的交换反应，产生重组后的 DNA 产物（框 11-2 图 2）。



框 11-2 图 2　FtsK 对染色体分离的调节。在细胞分裂前，图中所示的绿色的新复制起始位点移到细胞的两极，而以三角形表示包含 *dif* 的复制终点则停留在细胞中部。当 *dif* 位点复制完成后，两个子 *dif* 位点发生重组形成 Holliday 联结体，该联结体被 XerC 解离。如果复制后的染色体为单体形式，分离使突触复合体断开，*dif* 位点在细胞分裂前从细胞中部被移开；相反，如果染色体为二聚体（右侧图例），突触复合体仍被限制在细胞中央，位于细胞分裂位点的 FtsK 可以与之结合。FtsK 随后激活 XerD。XerD 介导了重组发生，然后是 XerC 介导的重组过程，最后二聚体解离成单体，完成细胞分裂。（来源：Barre F. X. et al. 2001. Proc. Natl. Acad. Sci. 98：8189-8195. Fig. 5，© National Academy of Sciences.）

FtsK是一个沿DNA移动的ATP酶，是一个功能类似于RuvB蛋白的“DNA动力蛋白机器”（DNA-pumping protein machine）。RuvB蛋白在同源重组过程中促进DNA的分支移位（在第10章讨论过）。FtsK还是一个膜结合蛋白，它被定位于细胞中发生细胞分裂的地方，其作用是在细胞分裂前将DNA从细胞中央移开，使细胞可以在这一位置分裂（框11-2图2）。

FtsK相对分裂位点的位置很关键，它确保了当细胞中出现二聚染色体时XerD被特异性地激活。这种情况下，染色体在分裂细胞的中央被“卡住”，二聚染色体的一半被分到每个子细胞中。FtsK也会不对称排列于dif位点、附近的特异极DNA（叫作KOPS）相互作用。结果，Ftsk运送dif位点趋近隔膜并互相接近。这一运动从而促进了dif位点的配对并激活XerD重组。就这样，在细胞分裂周期中，位点特异性重组被调控发生在合适的时间和地点上。

还有其他一些机制能够指导特定DNA片段的重组

尽管我们的讨论仅限于保守性位点特异性重组，实际上还有一些在特定序列上实现类似生物学功能的其他重组事件，其中有在第10章讲过的由于目标基因转换造成的酵母的交配型转换现象；在脊椎动物免疫系统中，编码关键蛋白质的基因片段在组装时发生的重排现象，即V(D)J重组，也是发生在特定的位点上。而后者的机制与转座很相似，所以在本章后半部分会对其作一些介绍。

转座

一些遗传因子通过转座被移位到染色体的新位置上

转座是一种特殊的遗传重组，它是将特定的遗传因子从DNA上的一个地方移位到另一个地方。这些可以移动的遗传因子叫做转座因子（transposable element）或转座子（transposon）。移位的发生是通过转座因子两端的DNA序列和宿主细胞DNA的一段序列间发生重组实现的（图11-16）。我们将看到，移位过程可能会也可能不会伴随着遗传因子的复制，某些情况下，这种重组过程会涉及暂时性的RNA中间体。

当转座因子移位时，转座因子通常对插入位点几乎没有序列选择性，所以，转座子可能会插入到基因内部，造成基因完全失活；也会插入到一个基因的调控序列中，造成该基因表达的改变。正是这种对基因功能和表达的影响使得转座子被发现（框11-4，及本章后面关于玉米因子和转座子的发现）。所以，并不令人意外的是转座因子是许多物种中新的突变的主要来源，而它们也的确是产生导致人类遗传疾病的突变的一个重要原因。正是转座因子这种不加选择的DNA插入能力，使得它们在经过修饰后被用于生物实验中，诱导突变和作为运输DNA的载体。

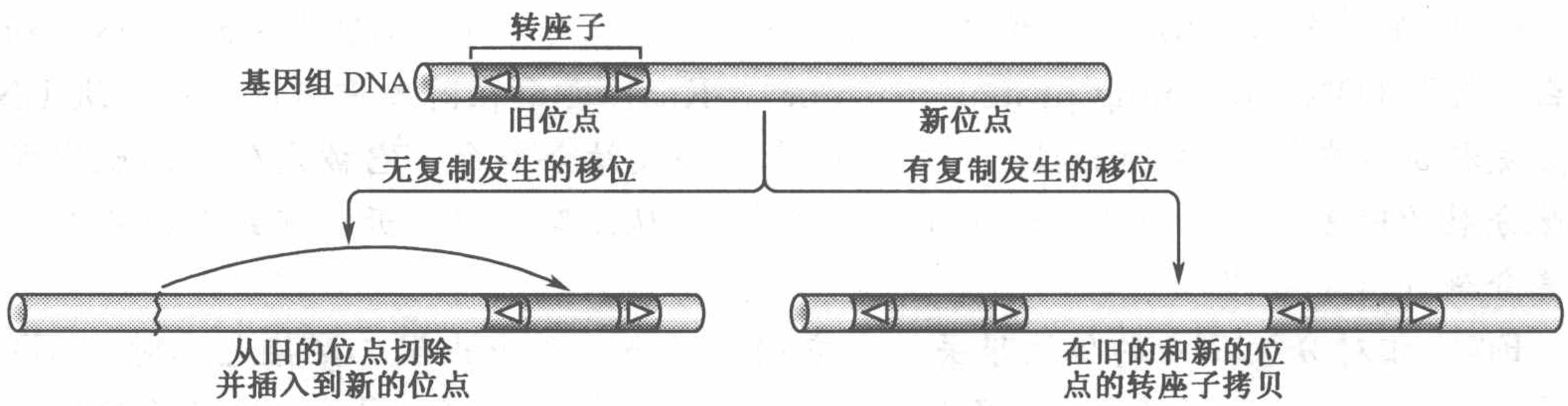


图 11-16 一个移动遗传因子转座到宿主 DNA 的一个新位点。某些情况下，重组过程涉及转座子从 DNA 上的旧位置被切除（左边）；在另一些情况下，转座子的一个拷贝保留在原来的位置上，另一个拷贝则插入到 DNA 的新位点（右边）。

所有生物的染色体上都有转座因子。基因组序列的比较分析揭示了两个惊人的发现：第一，转座相关序列可以占据物种基因组中的很大一部分，例如，人和玉米基因组的 50％以上是由转座相关序列组成的（包括转座子片段和被突变导致失活的“死”因子片段），而这与基因组中仅有很小一部分序列（在人的基因组中小于 2％）编码蛋白形成了鲜明的对比；第二，转座子的含量在不同的基因组中变化很大（图 11-17），例如，与人和玉米不同，果蝇和酵母的基因组则富含基因而缺乏转座子。

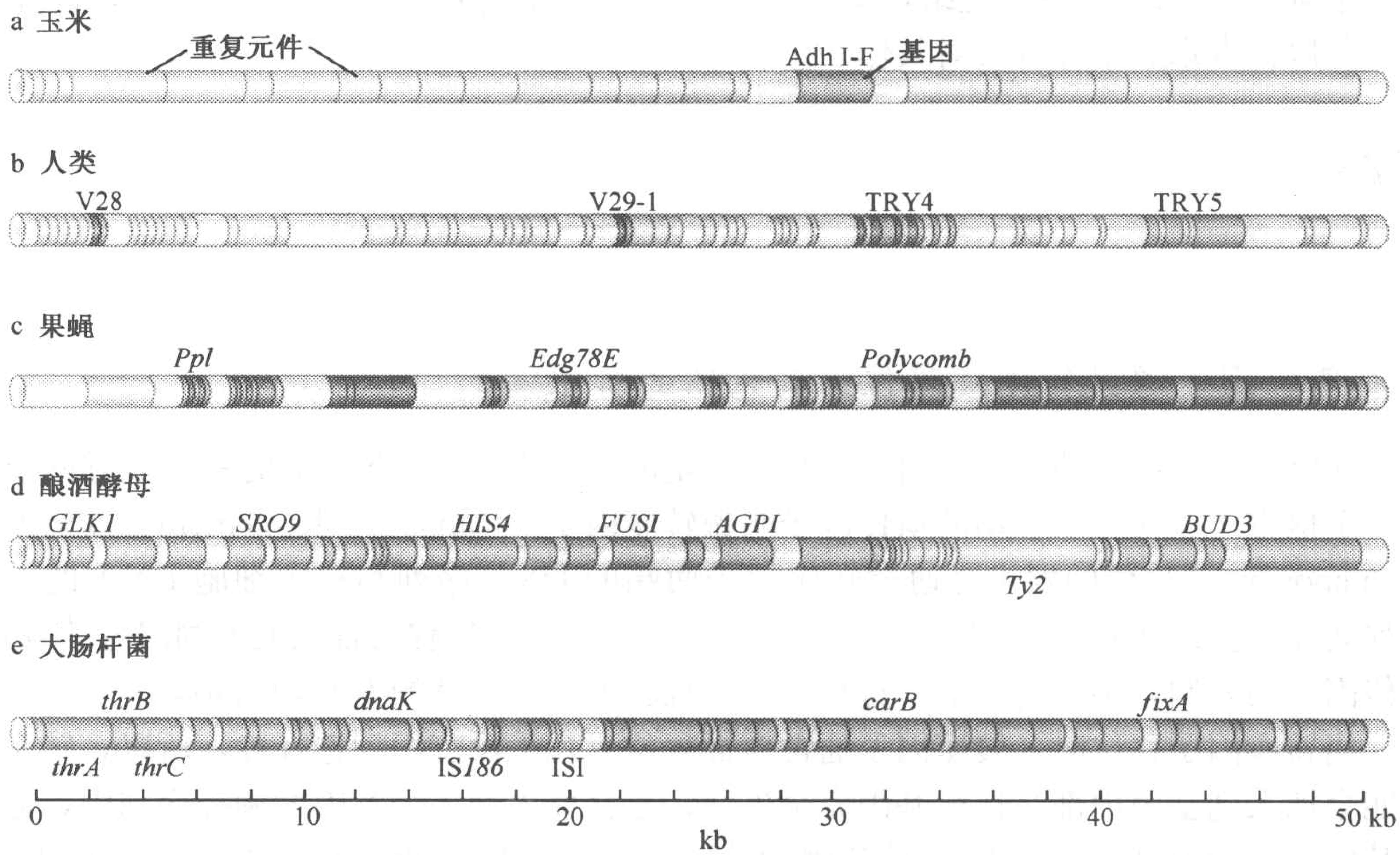


图 11-17 基因组上转座子的发生和分布。重复序列大多数由转座子或转座子相关序列所组成（如被截短的因子），以绿色表示，基因以紫色表示。a. 玉米；b. 人；c. 果蝇；d. 芽殖酵母；e. 大肠杆菌。（来源：Brown T. A. 2002. Genomes，2nd edition，p. 34，Fig. 2. 2 and references therein.）

转座因子有很多不同的类型，根据不同的结构和重组机制可将它们分为若干家族。下面我们介绍 3 种主要的转座子家族以及它们的重组机制，然后介绍几个被详细研究过的转座因子。在具体介绍转座因子时，我们将着重说明转座如何被调控以平衡它们的存在、复制与对宿主基因造成潜在的干扰和错误调控。

转座的遗传重组机制并不仅限于实现转座子的移位功能。例如，很多病毒在感染时，也是使用了一种与转座几乎相同的途径将自己整合到宿主细胞的染色体上，所以病毒的整合反应在这里与转座一同讨论。同样地，在细胞中改变基因表达模式的一些 DNA 重排也采用了一种类似于转座的机制。V(D)J 重组是一个被了解得很清楚的例子。它是脊椎动物免疫系统功能发育中一个必备的反应。V(D)J 重组将在本章最后进行讨论。

三类基本的转座因子

根据转座子的结构及其转座机制，可将其分为以下 3 个家族：

1. DNA 转座子。

2. 类病毒反转录转座子，包括反转录病毒。这些因子又称为 LTR（long temninal repeat ）反转录转座子。

3. 聚腺苷酸（poly-A）反转录转座子。这类因子又称为非病毒反转录转座子。

图 11-18 图解了这 3 类转座因子家族的基本遗传结构。DNA 转座子在整个重组周期中一直保持 DNA 的形态，它通过 DNA 链的断裂和重新连接实现移位，在这一点上它与保守性位点特异性重组因子移位的方式很相似。两种反转录转座子的移位都需要一个暂时性的 RNA 中间体的参与。

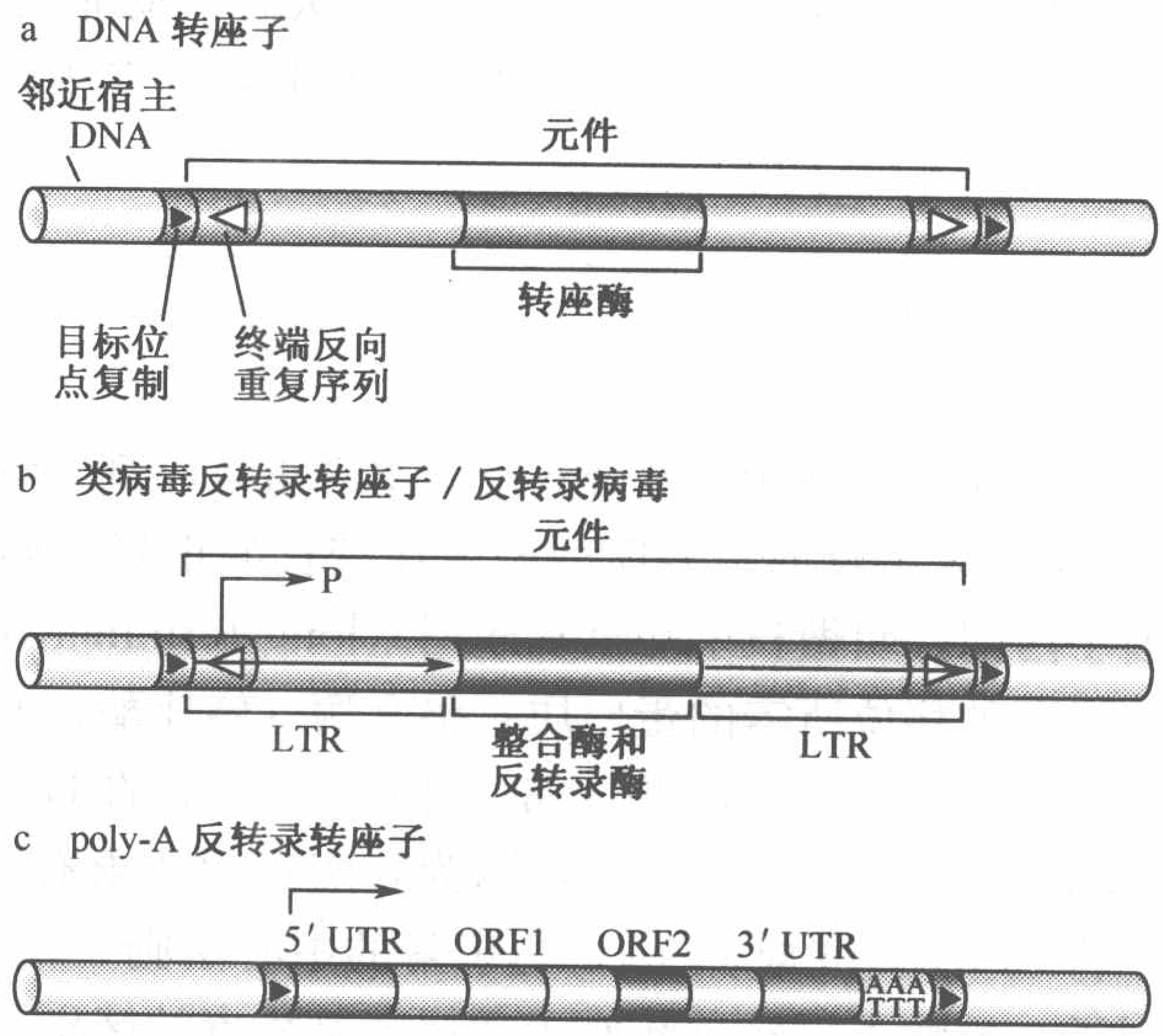


图 11-18 3 类转座因子的遗传结构。a. DNA 转座。其成分包括作为重组位点的终端反向重复序列（绿色含白色箭头）和一个编码转座酶的基因。b. 类病毒反转录转座子及反转录病毒。包括两端的长终端重复序列（LTR）及其间编码整合酶及反转录酶（RT）的区域。c. 聚腺苷酸反转录转座子。两端在 5′和 3′UTR 序列内，以及编码两个酶的序列：RNA 结合酶（ORF1）和一个同时具有反转录酶和核酸内切酶功能的酶（ORF2）。

DNA 转座子上带有一个转座酶基因，两端是重组位点

DNA 转座子上携带有既作为重组位点的 DNA 序列又编码参与重组的蛋白质的基因（图 11-18a）。转座因子两端的重组位点以反向重复序列排布，这些末端反向重复序列的长度从 25 到几百碱基对不等，它们并不是严格意义上的重复序列，并且带有重组酶识别序列。催化转座的重组酶通常叫做转座酶（transposase）[有时也叫整合酶（integrase）]。

DNA 转座子带有编码自身转座酶的基因，也可能会携带一些其他基因，这些基因编码的蛋白质可能调节转座，也可能为转座因子或宿主细胞提供一些有益的功能。例如，很多细菌 DNA 转座子上携带的基因所编码的蛋白质能够提高细菌对一个或多个抗生素的抗性。因此，这种转座子的存在使得宿主细胞具有了对该抗生素的抗性。

紧邻转座子两端的 DNA 序列有一个短的重复序列片段（2～20bp），这些片段的排布为正向重复，叫做靶位点重复（target site duplication），产生于重组过程中，我们将在下面对此进行讨论。

转座子包括自主转座子和非自主转座子

带有一对末端反向重复序列和一个转座酶基因的 DNA 转座子，本身具备了促成自身转座的所有条件，这类因子叫做自主转座因子（autonomous transposon）。然而，在基因组上还有很多比较简单的可移位 DNA 片段，叫做非自主转座子（nonautonomous transposon），它们仅有末端反向重复序列，即转座所需的 *cis*-acting 序列。细胞中自主转座子表达的转座酶会识别这些末端反向重复序列，从而帮助非自主转座因子发生移位。但是在没有“帮助者”转座子（产生转座酶）存在时，非自主转座因子只能待在原处无法移动。

类病毒反转录转座子和反转录病毒带有末端重复序列和两个对重组很重要的基因

类病毒反转录转座子和反转录病毒也带有反向末端重复序列，它们是重组酶结合和作用的位点（图 11-18b）。末端反向重复序列嵌入在较长的重复序列中，它们以正向重复形式排列在转座因子两端，叫做长末端重复序列（long terminal repeat）或 LTR。类病毒反转录转座子编码两种移位所需的蛋白质：整合酶（转座酶）和反转录酶。

反转录酶（reverse transcriptase，RT）是一种特殊类型的 DNA 聚合酶，它能以 RNA 为模板合成 DNA。转座需要有这个酶，因为转座过程要求有一个 RNA 中间体的参与。由于这些因子将 RNA 转变成 DNA，与正常的中心法则（DNA 到 RNA）相反，所以称为“反式”因子。类病毒反转录转座子和反转录病毒的区别是：反转录病毒的基因组被组装到病毒颗粒内，离开宿主细胞再去感染新的细胞；而反转录转座子只能在一个细胞内移位到一个新的 DNA 位点，但不能离开宿主细胞。与 DNA 转座子一样，这些因子两端也是短的靶位点重复序列，它们是在重组过程中产生的。

像基因的聚腺苷酸反转录转座子

聚腺苷酸反转录转座子没有其他转座子所带的末端反向重复序列，相反，它的两端序列含有完全不同的成分（图 11-18c）。一端称为 5′UTR，另一端是3′UTR，它带有一串叫做聚腺苷酸序列（poly-A sequence）的 A-T 碱基对，这些因子两端也有短的靶位点重复序列。

反转录转座子带有两个基因，ORF1 和 ORF2。ORF1 编码一个 RNA 结合蛋白，ORF2 编码的蛋白质同时具有反转录酶和核酸内切酶的作用。尽管这个蛋白质与其他转座子家族编码的转座酶和重组酶不同，但它在重组过程中也起着关键作用。与 DNA 转座子和类病毒转座子一样，聚腺苷酸反转录转座子也存在自主转座子和非自主转座子两种形式。此外，基因组序列分析表明，很多 5′UTR 序列不完整的断裂转座子失去了转座的能力。

DNA 转座的剪切-粘贴机制

DNA 转座子、类病毒反转录转座子和反转录病毒都用一种类似的重组机制把它们的 DNA 插入到一个新的位点。我们首先来了解一下最简单的转座反应：DNA 转座子非复制性移位的机制。这个重组过程包括转座子从 DNA 上初始位置的切除以及将这个切下来的转座子整合到一个新的 DNA 位点，因此把这个机制叫做剪切-粘贴转座（cut-and-paste transposition）（图 11-19）。

开始重组时，转座酶结合到转座子两端的反向重复序列上。一旦转座酶识别了这些序列，它就将这两端聚集在一起形成一个稳定的蛋白质-DNA 复合体，这个复合体被称为突触复合体（synaptic complex）或转座体（transpososome），它包含一个转座酶的多聚体（通常为 2～4 个亚基）和两个 DNA 末端（见下文）。这个复合体的功能是确保转座子移位所需的 DNA 断裂和重新连接能够在转座因子两端同时发生，此外，还能在重组过程中保护 DNA 末端免受细胞中酶的作用。下一步是将转座子 DNA 从基因组的原始位置上切除。转座子上的转座酶亚基先断裂转座子每端的一条单链，断裂位置就在转座子 DNA 和它所插入的宿主序列间的连接处（该区域叫做 flanking host DNA）。转座酶切断 DNA 后，转座因子 DNA 末端带有游离的 3′羟基基团。要完成断裂反应，每端的另一条 DNA 单链也需要被切断，而不同的转座子断裂这第两条单链的机制各不相同（这些单链的 5′端位于转座子与宿主 DNA 连接处），这些机制将在下面将进行描述。

转座子被切除后，被转座酶首先释放的 3′端羟基攻击新插入位点的 DNA 磷酸二酯键，被攻击的 DNA 片段叫靶 DNA（target DNA）。回想对大多数转座子而言，靶 DNA 本质上可以是任何序列。攻击的结果是将转座子 DNA 共价连接到靶 DNA 上，而在每个 DNA 连接反应中，靶 DNA 上会产生一个缺口（图 11-19）。DNA 连接反应是通过一步酯基转移作用实现的，它又被称为 DNA 单链交换（DNA strand transfer）。（图 11-20）RNA 剪接中就使用了类似的机制来连接核酸单链（第 13 章）。

转座体确保转座子两端的 DNA 一起攻击同一靶位点的两条链，这两个链上被攻击的位点通常相差几个碱基（如比较常见的是相隔 2 个、5 个和 9 个碱基），这个距离对

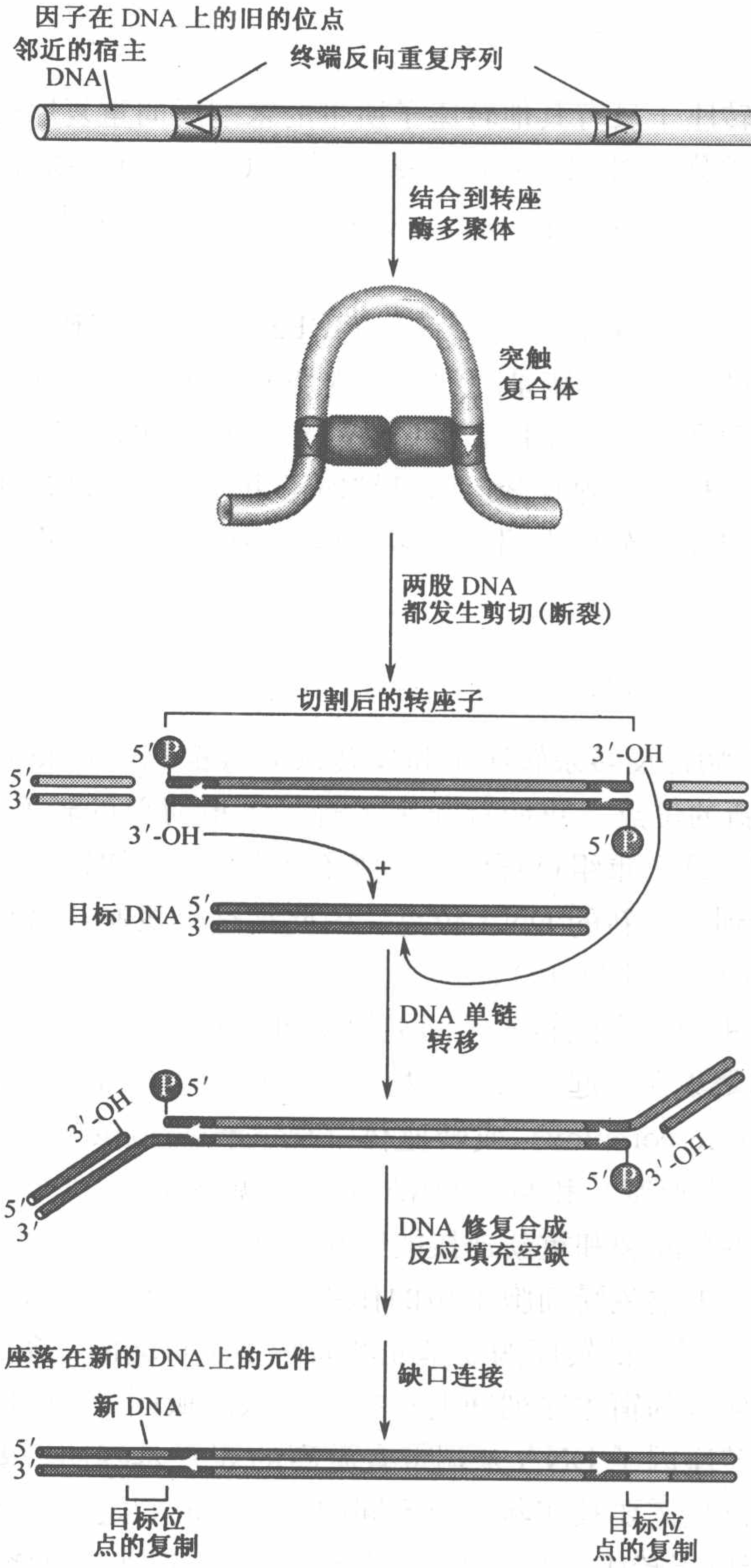


图 11-19　**剪切-粘贴的转座机制。**如图所示，一个转座子从灰色宿主 DNA 的一个靶位点转移到蓝色 DNA 上的新位点。注意在 DNA 单链交换反应中靶 DNA 断裂位点的黏性末端使新的靶位点上生成短的重复序列（靶位点复制）。原插入位点的 DNA（灰色）由于转座子的切除留下了一个双链的断裂，这个断裂可以通过非同源末端连接或同源重组的方式被修复（见第 9 章和第 10 章）。注意：为简洁起见，图中的化学步骤未标示出结合蛋白质。

每种转座子来说都是固定的，由此产生邻近转座子两端的靶位点短重复序列（将在下节说明）。一旦 DNA 单链转移完成，转座体的任务也就完成了，剩下的重组步骤由细胞内的 DNA 修复蛋白来完成。

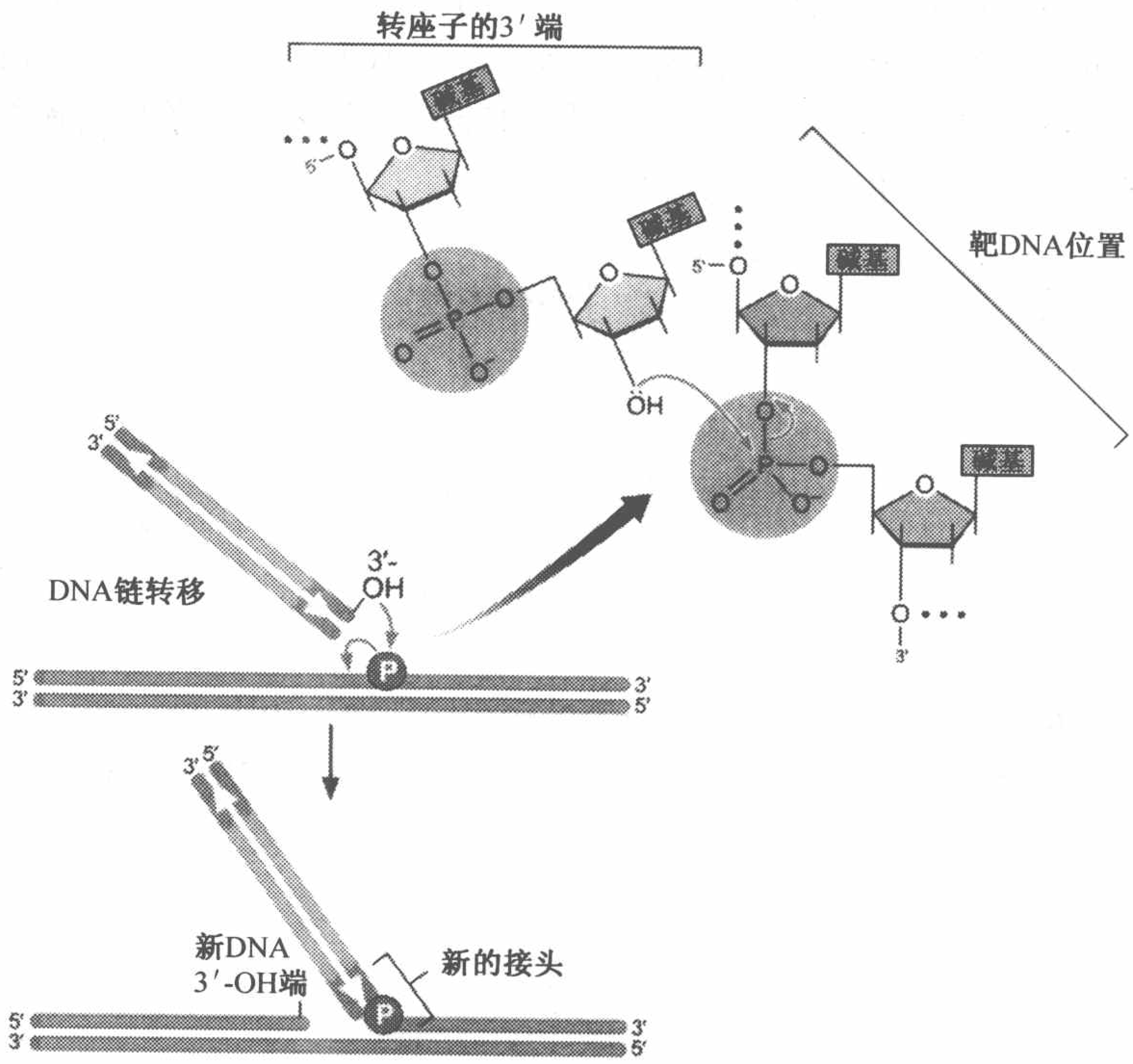


图 11-20 DNA 单链转移的详细化学步骤为简洁起见，图仅示转座子和靶 DNA 的单链。

剪切-粘贴转座中的中间体由缺口修复来完成

在 DNA 单链转移后产生的 DNA 中间体结构使转座子 DNA 的 3′端结合到靶 DNA 上。这个结构还带有靶 DNA 上的两个缺口，它们产生于 DNA 单链转移过程中。转座子 DNA 单链上的两个位点被转移到靶位点的两条单链上，这两条单链的断点相隔几个核苷酸，这导致在插入的转座子两端出现了短小的单链 DNA 缺口，这些缺口会被宿主细胞自身编码的 DNA 修复聚合酶填补。要说明的是，DNA 单链转移步骤中靶 DNA 断裂所产生的 3′羟基端可以作为修复合成中的引物（图 11-18)。填补缺口的结果是造成产生邻近转座子的靶位点的复制（见上面)。所以，靶位点重复序列的长度表明了 DNA 单链转移中靶 DNA 的两条单链上被攻击位点的距离。在缺口的修复完成后，还需要 DNA 连接酶来封闭 DNA 单链。

剪切-粘贴转座在它原来插入的 DNA 位点处也留下了一个断裂的双链，它必须被修复以维持宿主细胞基因组的完整。通过同源重组来修复双链 DNA 断裂在第 10 章中已讲过，而这些断裂有时也会直接重新连接起来。下面在 Tc*1*/*mariner* 转座子家族的讨论中我们就会看到这个现象。

DNA 转座中断裂非转移链的多种机制

如前所述，转座酶通过断裂转座因子的 3′端和促进 DNA 单链的转移来实现剪切-

粘贴转座过程。然而，通过这种机制移动的转座子也需要断开转座子与邻近宿主 DNA 连接处的 5′端链。这些单链叫非转移链（nontransferred strand），其 5′端在 DNA 单链转移反应期间并不直接连到靶 DNA 上。不同的转座子催化第二条单链断裂反应的机制各不相同（图 11-21），这里介绍 2 种方法。

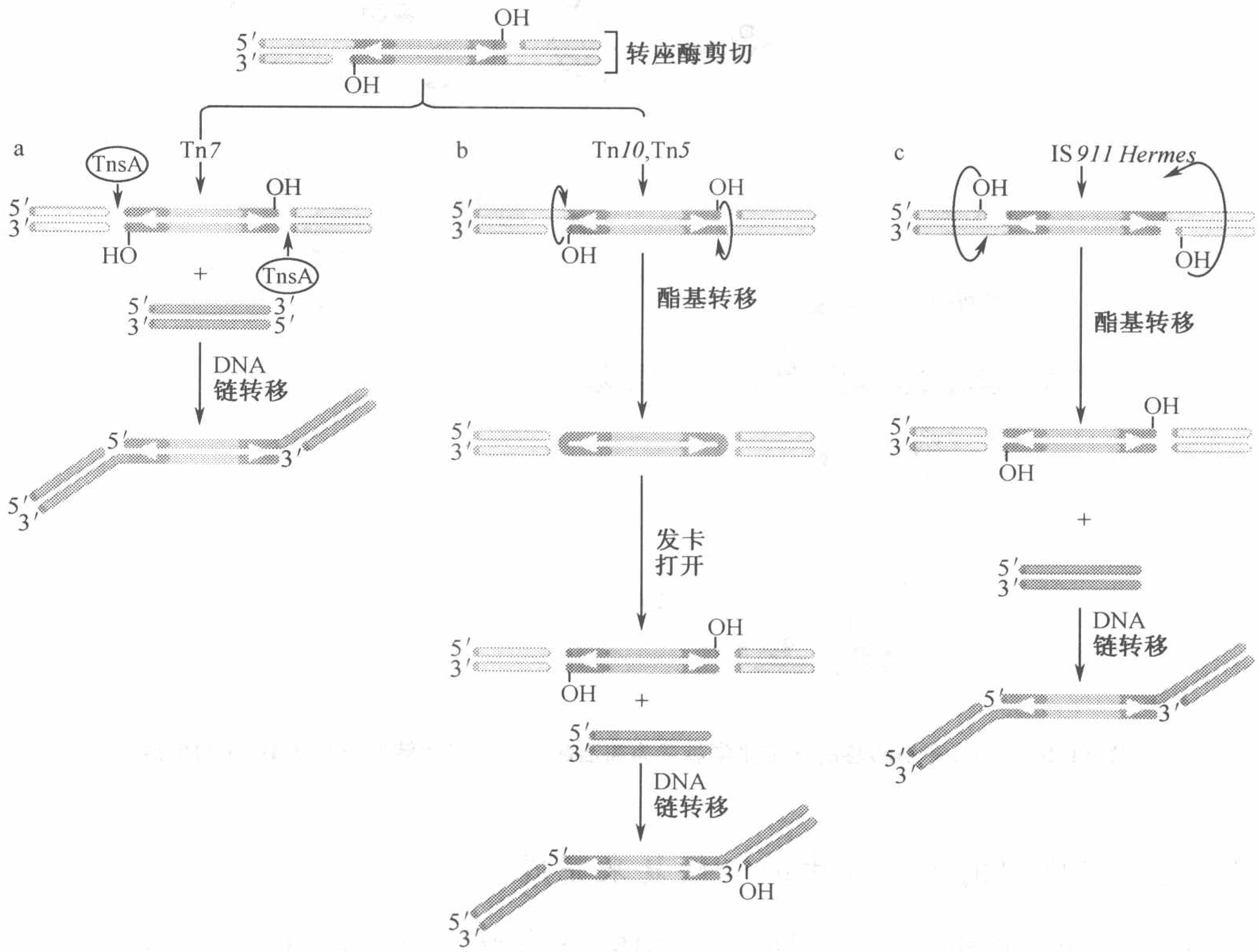


图 11-21 非转移链断裂的 3 种机制。 a. 通过一种非转座酶来切断。b. 转座酶催化一条单链攻击相对链形成发卡形的 DNA 中间体，在这种机制中，是转移链攻击非转移链。然后这两个发卡形末端再被转座酶水解打开。c. Hermes 转座子使用的是通过形成发卡切割第二条单链的另一种机制。在这种机制中，先导链（非转移链）首先断裂，并且发卡结构在插入起始位点形成，而不是在转座子末端。

用来断裂非转移单链是非转座酶的一种酶（图 11-21）。例如，细菌的 Tn*7* 转座子编码的一种特殊蛋白质（TnsA）就有这种作用（图 11-21a）。TnsA 的结构与核酸内切酶很相似，它与 Tn*7* 自身编码的转座酶（TnsB 蛋白）组装在一起，共同作用将转座子从原来的位点上切除下来。

断裂非转移链的另一种方法由转座酶自身完成，它用一种 DNA 酯基转移机制，其过程与 DNA 单链转移相似。例如，在 Tn*5* 和 Tn*10* 转座子断裂非转移链时就产生一个“DNA 发卡”结构。为了形成发卡结构，转座酶用先断开的转座子 DNA 3′端跨过 DNA 双螺旋直接攻击另一条单链的磷酸二酯键（图 11-21b），结果既断裂了被攻击的 DNA 单链，又将其与转座子 DNA3′端与缺口的一端共价连到一起。结果，两条 DNA 单链通过一个环形末端共价连接在一起，其结构让人联想到一个发卡。

然后发卡 DNA 的末端被转座酶断裂（打开），产生一个标准的双链 DNA 断裂，打开反应在转座子 DNA 两端进行。这一步骤一旦完成，DNA 因子的 3′端就准备通过 DNA 单链转移反应连接到一个新的靶 DNA 上。

Hermes 转座子是 *hAT* 因子家族的一个成员，也通过 DNA 发卡中间体从旧的 DAN 插入位点上切除转座子，但是切割单链和酯基转移反应的顺序有所不同。这里 DNA 发卡在宿主细胞的 DNA 上形成，而不是在转座因子的末端（图 11-21c）。在本章的后面我们会看到，这种 DNA 切割和重新连接的方式与研究发现的 V(D)J 重组的早期步骤是十分相似的。这种机制的相似性成为一个假说的有力支持，该假说认为，V(D)J 重组是在脊椎动物进化过程中，由宿主机体捕获和“驯服”转座子形成的。

尽管在图 11-21 中未标示出来，转座子两端 DNA 的断裂也可以通过转座子两末端间的酯基转移反应完成。在这种情况下，一个断裂的 3′-OH 端攻击转座因子上同一单链的另一端，产生的 DNA 中间体进一步反应得到切除的转座子。IS*3* 转座子家族就是使用这种机制。

为什么转座酶会用酯基转移作用作为断裂机制呢？这可能是一种经济的解决办法。转座酶具有以下基本功能：① 催化转座子 DNA 3′端的位点特异性水解；② 用酯基转移作用将此末端变为非特异性位点。而当同样的反应作用于一个新的 DNA 位点时，(例如：起始切割位点的相对链）就可以帮助转座酶将转座子切除，因此，这个机制就使转座子无需再编码一个酶来断裂非转移链。

DNA 的复制性转座

一些 DNA 转座子使用的机制是复制性转座（replicative transposition），其转座因子在每个转座过程中进行了复制。虽然转座反应的结果明显不同，但我们可以看出，其重组机制与剪切-粘贴转座非常相似（图 11-22）。

复制性转座的第一步是转座酶蛋白和转座子 DNA 两端装配成一个转座体。像在剪切-粘贴转座中看到的一样，形成转座体对协调转座子 DNA 两端的断裂和重新连接是非常重要的。

第二步是转座子末端 DNA 的断裂。这个反应是在转座体内由转座酶催化的。转座酶在转座子序列与邻近的宿主 DNA 连接处诱导产生一个缺口（图 11-22），这个断裂使得转座子序列的两个 3′-OH DNA 末端获得释放。与剪切-粘贴转座不同，这时转座子 DNA 并不从原来的序列上被切除，这一点是这两种转座最大的不同。

然后，转座子 DNA 的 3′-OH 末端通过 DNA 单链的转移反应被连接到靶 DNA 位点上，这一机制与前面的剪切-粘贴转座相同。但是，这里产生的中间体是双交叉的 DNA 分子（图 11-22）。在此中间体中，转座子的 3′端被共价连接到新的靶位点，而 5′端仍与原来的邻近 DNA 相连。

中间产物的两个 DNA 交叉形成一个复制叉结构（第 8 章）。DNA 单链转移后，宿主细胞的 DNA 复制蛋白将组装到交叉上。在一个已经充分研究的复制性转座例子中（Mu 噬菌体，下面将讲到），蛋白质特异性地组装到两个交叉中的一个上（图 11-22 下

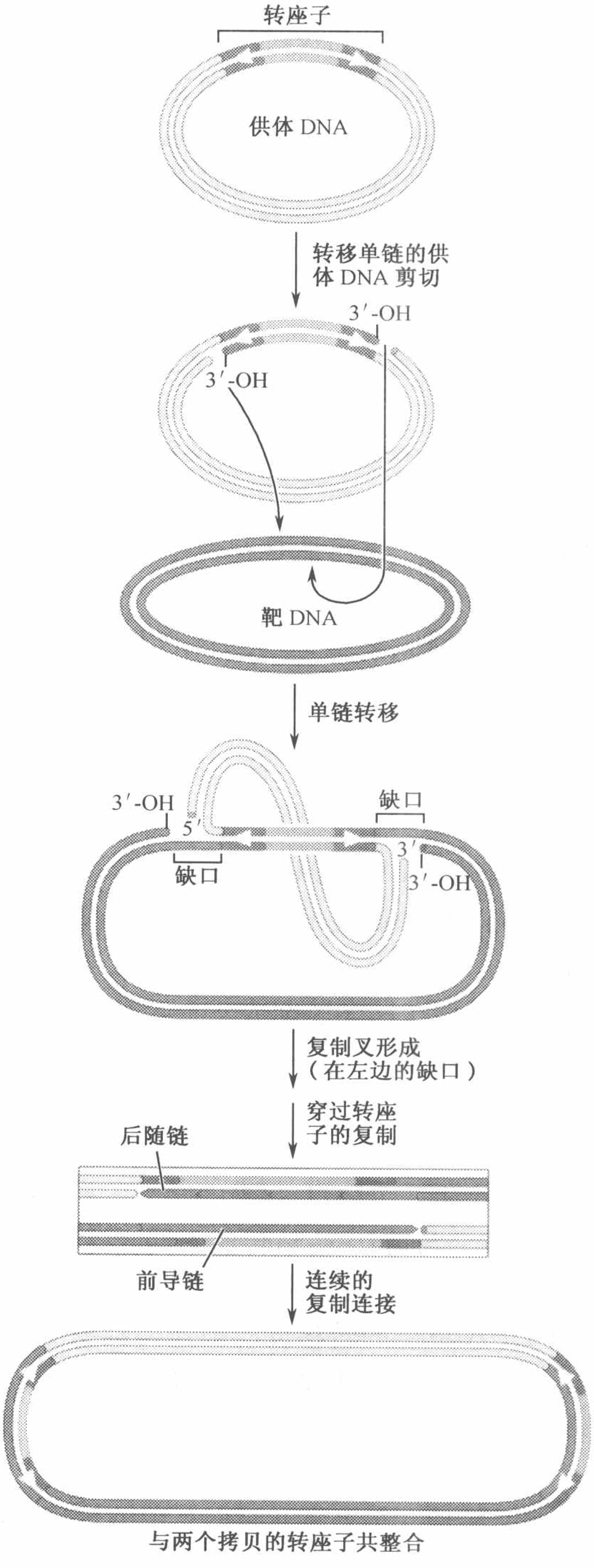


图 11-22 复制性转座的机制。转座体在转座子两端诱导形成缺口，露出 3′-OH 端。然后这些羟基攻击靶 DNA，并通过 DNA 单链转移与之连接。注意在转座子的每一端只有一条单链插入靶位点，形成了一个双交叉的 DNA 结构。细胞的复制成分组装到其中一个“交叉”（图中左边），将整个转座子序列进行复制。生成的产物叫做共和体，它有两个通过转座子两个拷贝连接起来的起始环形 DNA 分子。交叉中间体的单链 DNA 缺口产生了靶位点的复制，这些复制在图中没有明确标出。

图)。以断开的靶 DNA 3′端为引物进行 DNA 合成，复制沿转座子序列进行直到第二个交叉处停止。此复制反应产生 2 个拷贝的转座子 DNA，复制产物两端是短小的正向排列靶位点重复序列。

复制性转座常常会造成染色体的倒位和丢失，这对宿主细胞有很大的危害性。这种趋势所导致的重排有时会对复制性转座造成不利的结果。或许这就是为什么很多转座因子进化出在插入到新位点前即将其从原位点完全切除的方法。转座子切除后就避免了对宿主基因组造成较大的干扰。下面我们将会讨论到，通过一个 RNA 中间体也可以避免转座造成的干扰性重排。

类病毒反转录转座子和反转录病毒通过一个 RNA 中间体实现移位

类病毒反转录转座子和反转录病毒插入到宿主细胞基因组的新位点，所用的 DNA 断裂及 DNA 单链转移步骤与前面讲的 DNA 转座相同。但不同的地方是，这些反转录因子的重组需要一个 RNA 中间体的参与。

转座周期开始于反转录转座子（或反转录病毒）DNA 序列在细胞内的 RNA 聚合酶的催化下转录得到 RNA。转录起始于 LTR（图 11-23）中的一个启动子序列，一直穿过整个因子得到转座因子的一个几乎全长的 RNA 拷贝，然后 RNA 反转录成一个双链 DNA 分子，这个 DNA 分子叫做 cDNA（复制的 DNA），它游离于宿主 DNA 序列之外。

这个 cDNA 能够被整合酶（与 DNA 因子的转座酶高度相关的一种蛋白质，下面将会看到）识别并被重组到一个新的靶 DNA 位点。整合酶结合到 cDNA 的末端，然后从每条单链的 3′端切下几个核苷酸，这个断裂反应与 DNA 转座子的 DNA 断裂步骤相同。当用于整合的直接前体 DNA 由 RNA 模板反转录生成后，它已经呈一个被切除的转座子形态。因此，对这些因子而言无须一个切断第二单链的机制。然后通过 DNA 单链转移反应，整合酶催化将断开的 3′端插入到宿主细胞基因组的靶位点。上面讲过，这些靶位点本质上可以是任何 DNA 序列，DNA 单链转移后在靶位点产生的缺口由宿主细胞的 DNA 修复蛋白填补并完成重组。这个缺口修复反应造成靶位点的复制。

由于产生 RNA 中间体的转录过程起始于 LTR 的内部，所以 RNA 没有携带全长的 LTR 序列，转录起始位点到转座因子末端之间的序列丢失了。因此，在反转录过程中需要一个特殊的机制来产生全长的转录因子。在反转录过程中，有两个内部引物合成和两个单链转换反应（本过程详见框 11-3，反转录病毒 cDNA 的合成途径）。这些转换造成了 cDNA 末端序列的复制，这样 cDNA 序列就完整了，重构的 LTR 序列补偿了转录过程中丢失的区域。LTR 的这次重构对于重组酶识别 cDNA 和后续的重组过程是至关重要的。

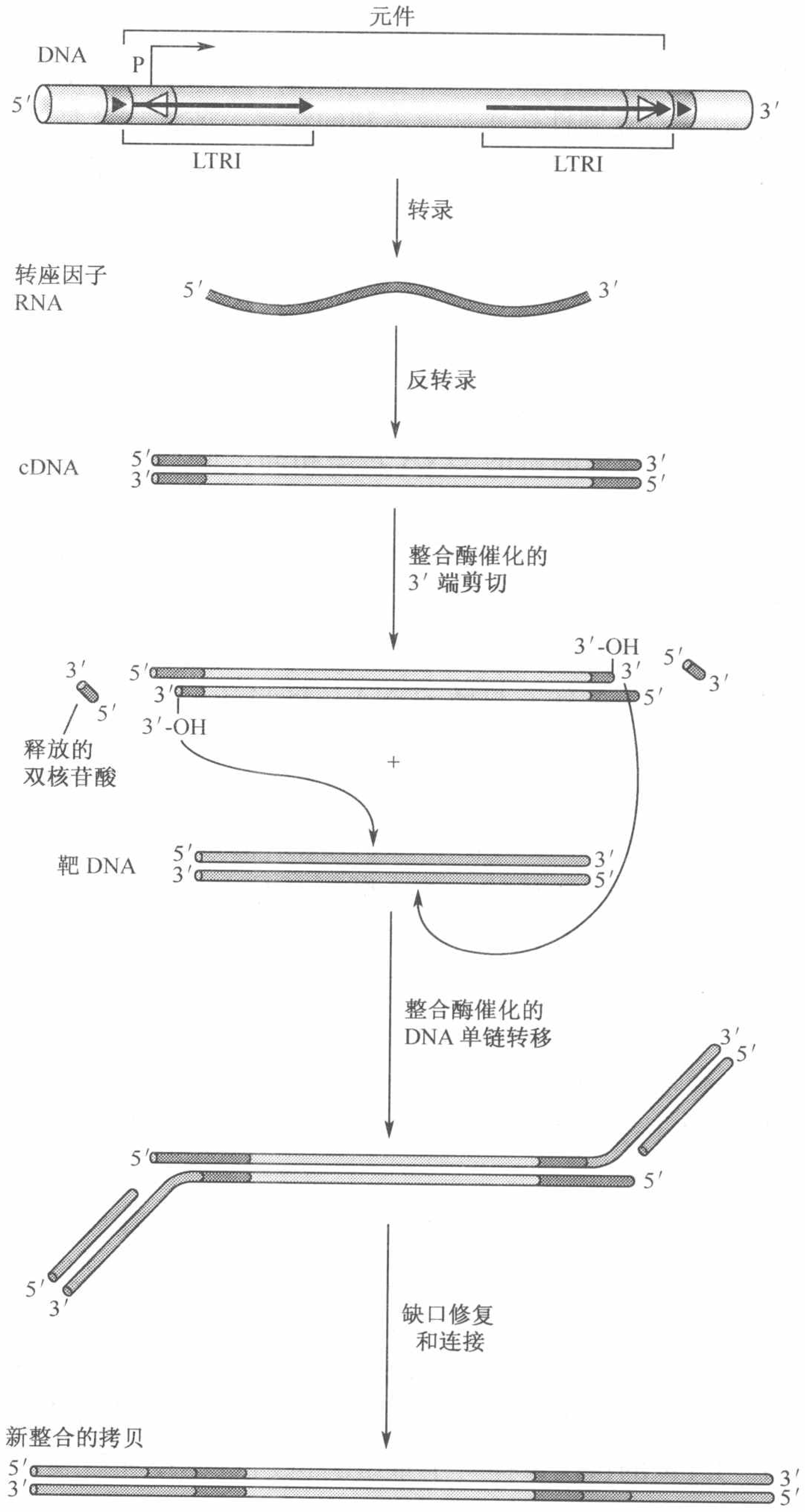
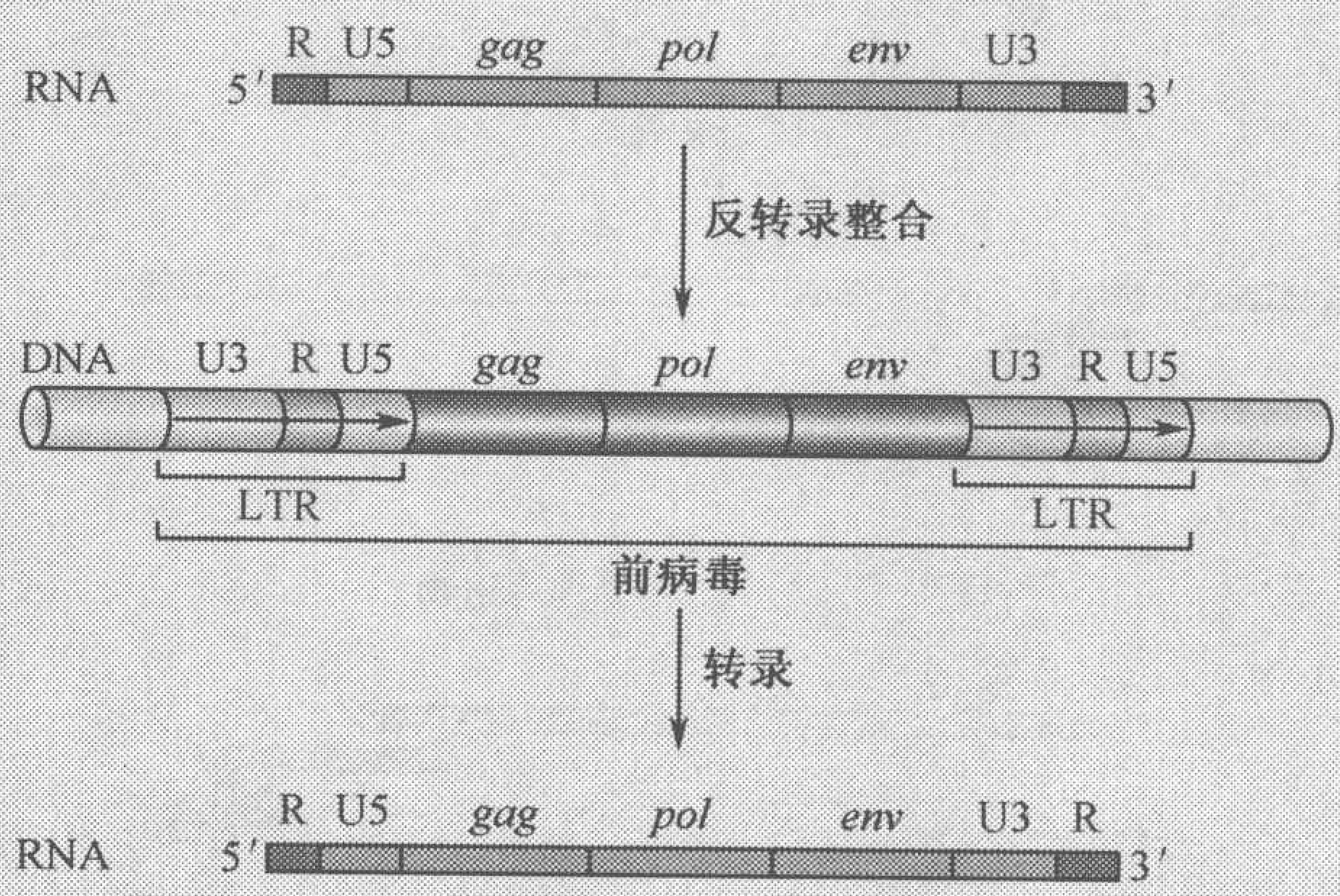


图 11-23 反转录病毒整合机制和类病毒反转录转座子的转座。上边图例表示已整合的前病毒。LTR 序列的详细图示见框 11-3。如图所示，病毒 RNA 的转录启动子位于左边的 LTR 中，该病毒 RNA 的 cDNA 合成过程详见框 11-3。整合酶催化的 DNA 剪切和单链转移步骤如此图所示。

框 11-3 反转录病毒 cDNA 的合成途径

要理解反转录病毒（或类病毒反转录转座子）的反转录过程，我们先要详细了解一下 LTR 序列的结构。每个 LTR 由 3 个序列因子构成：U3（独特的 3′端）、R（重复序列）和 U5（独特的 5′端）。反转录病毒基因组的整合拷贝转录生成了两端为 R 序列的病毒 RNA（框 11-3 图 1）。因此，在反转录过程中，必须要合成一个额外的 U3 和 U5 区。下面将要讲到，发生复制的原因是在 RNA 基因组内部引发了 DNA 的合成和在复制过程中 R 序列使得“单链转换”发生两次。



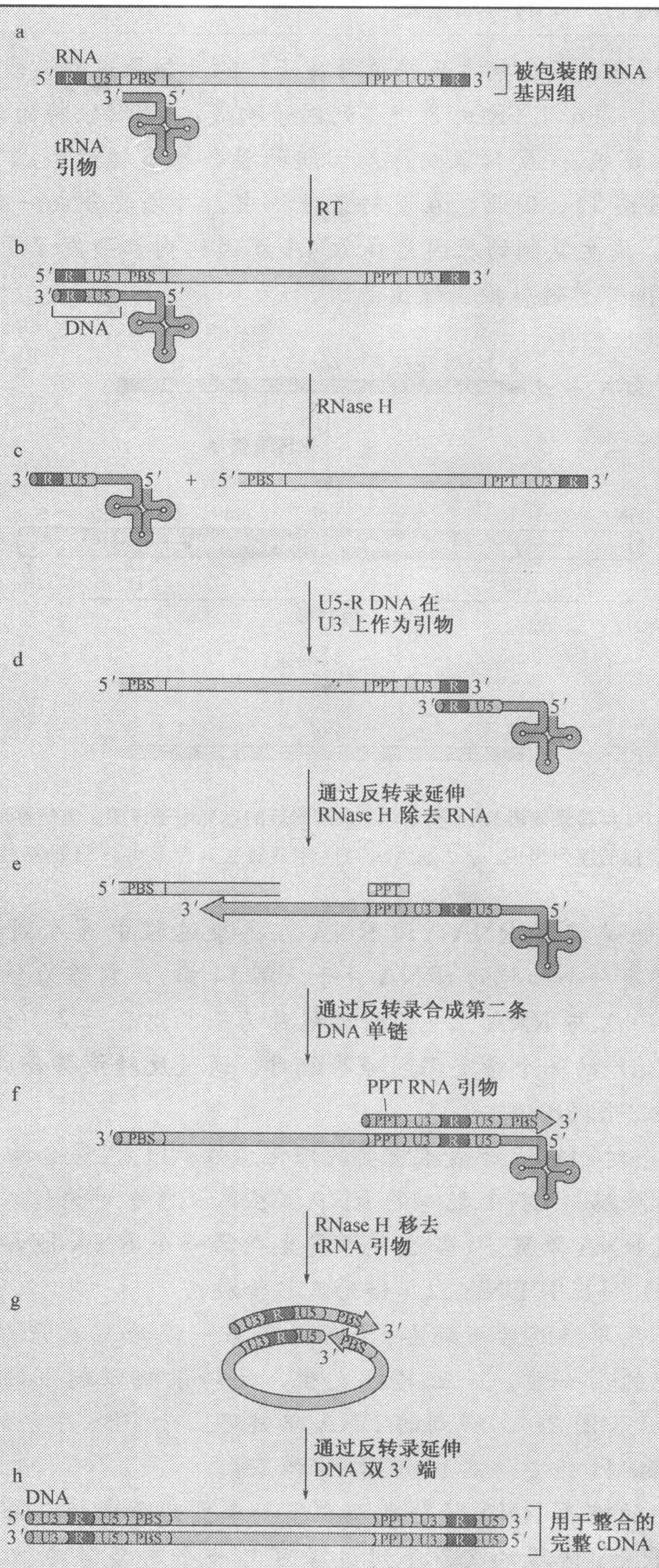
框 11-3 图 1　反转录病毒 RNA 和 cDNA 末端附近的序列因子详图。类病毒反转录转座子有十分类似的序列组成。*pol* 基因编码反转录酶（具有 RNase H 的活性）和整合酶。

装入病毒颗粒的是病毒 RNA，该 RNA 在感染过程中进入到新细胞中。与病毒 RNA 一起装入的还有一个细胞的 tRNA 分子（第 14 章），它作为合成第一条 cDNA 单链的引物。这个 tRNA 与 RNA 上 U5 附近被称为引物结合位点（primer-binding site, PBS）（框 11-3 图 2a）的一个特定序列碱基配对，通过反转录酶合成 U5 区和第一个 R 片段的拷贝（框 11-3 图 2b）。

反转录酶有两个对 cDNA 合成很重要的催化活性：DNA 聚合酶的活性和 RNase H 的活性。RNase H 降解与 DNA 配对的 RNA（在第 8 章中讨论过）。在反转录过程中，RNase H 移去模板 RNA 单链。（当这一步发生在第一个 RNA-DNA 杂交中间体上时，框 11-3 图 2b 和 2c），U5-R DNA 以单链形式被释放。

这条 U5-R DNA 单链随后与病毒 RNA 分子另一端的 R 区配对（框 11-3 图 2d），这是两条单链转换的第一步。一旦转换发生，反转录酶以剩下的 RNA 为模板继续 DNA 的合成（框 11-3 图 2e），得到的 DNA 单链终止于 PBS 序列的 3′端，这条 RNA 模板链同样被 RNase H 移去（框 11-3 图 2d 和 2e）。

RNase H 介导的病毒 RNA 降解也产生了一个作为合成第二条 cDNA 单链引物的 RNA 片段，这一 RNA 区域的 U3 序列边缘可与一个叫多嘌呤区（polypurine tract, PPT）保持碱基配对（框 11-3 图 2e 和 f）。这一引物的延伸将 U3、R、U5 和 PBS 序列复制入 DNA。



框 11-3 图 2 生成反转录病毒或反转录转座子 RNA cDNA 拷贝的反转录途径。

一旦 tRNA 引物从第一条 cDNA 单链上移除，第二单链的转换随即发生，3′端的PBS 互补序列促使两个 DNA 单链间配对和环形中间体的形成。中间体上每个 3′端向另一端的延伸产生了具有完整 LTR 序列的双链 cDNA，这时这个 DNA 分子就准备通过整合酶蛋白的作用整合到细胞基因组中。

反转录酶是由病毒编码（或反转录转座子编码）的酶，不具备任何重要的细胞功能。然而，其对反转录病毒的复制是至关重要的。所以它又是抗病毒药物的常见靶点，包括很多用于治疗艾滋病的药物。

DNA 转座酶和反转录整合酶同属于一个蛋白超家族

由上可以看出，在 DNA 转座和类病毒反转录转座子及反转录病毒的 DNA 移位过程中，转座子 DNA（或 cDNA）3′端的断裂和单链转移反应都是共有的步骤。这一重组机制的保守性可由转座酶及整合酶蛋白的结构得到反映（图 11-24）。高分辨率结构揭示出许多不同的转座酶和整合酶上的催化结构域有着相同的三维结构。这个催化结构域包含了 3 个进化保守的酸性氨基酸：两个天冬氨酸（D）和一个谷氨酸（E）。所以，这

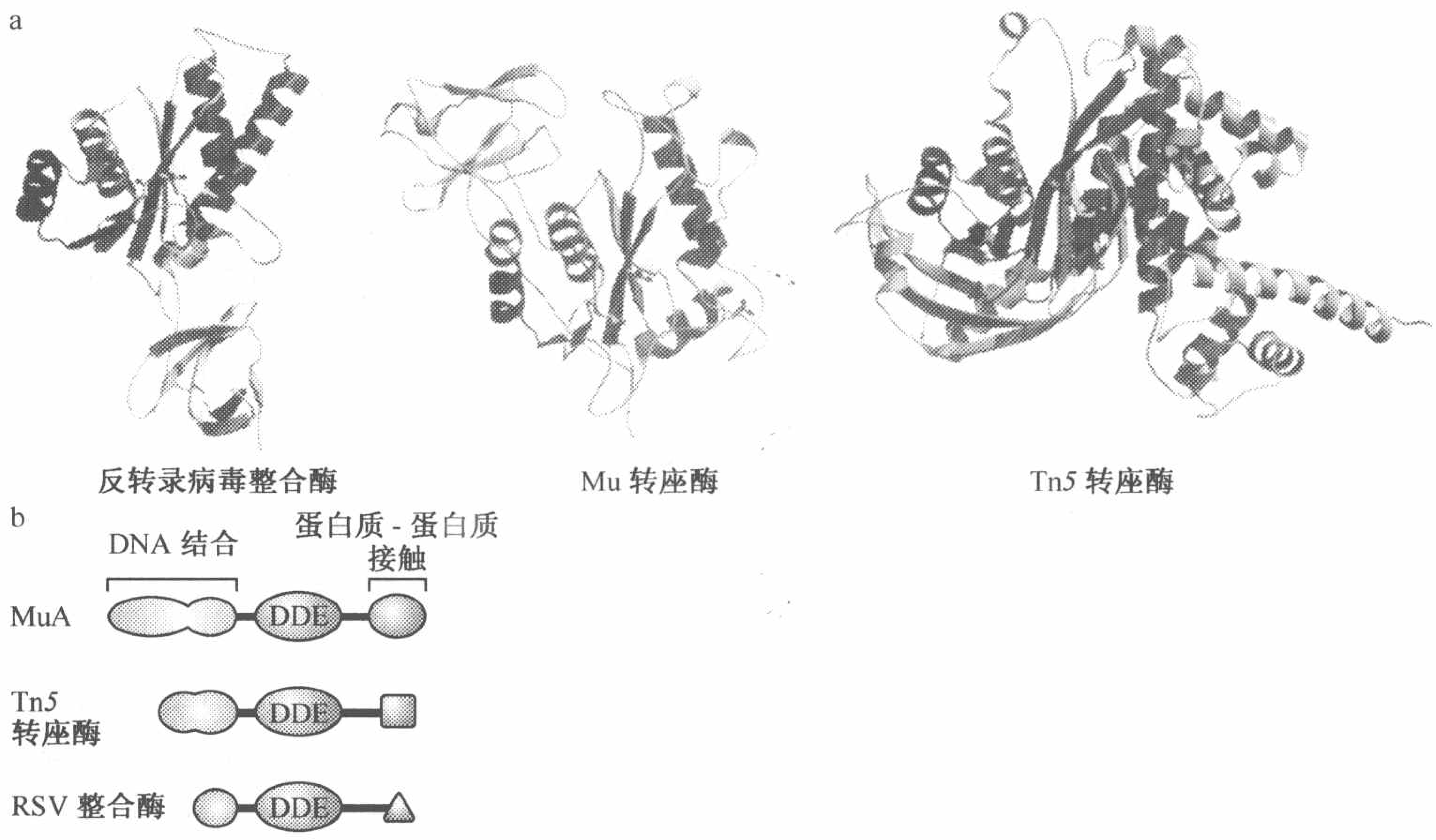


图 11-24 转座酶和整合酶催化结构域的相似性。 a. 从右至左依次为 Tn5 转座酶（Davies D. R. et al. 2000. Science 289：77-85）、Mu 噬菌体转座酶（Rice P. and Mizuuchi K. 1995. Cell 82：209-220）和 RSV 整合酶（Chook Y. M. et al. 1994. J. Mol. Biol. 240：476-500）的保守核心结构域。它们相同的二级结构成分以相同的颜色表示。DDE 基序（motif）的活性位点残基以球棒模型表示。图形通过 BobScript、MolScript 和 Raster 3D 软件制作。b. 如 a 中所示的 3 个蛋白质的结构域的图解，N 端的域结合 DNA，图 a 的中间的结构域包含了催化区，C 端的结构域参与组装转座体所需的蛋白质与蛋白质相互作用，以及与调节重组的其他蛋白质发生相互作用。（来源：Rice P. A. and Baker T. A. 2001. Nat. Struct. Biol. 8：302-307.）

类重组酶被归类为 DDE 基序（motif）的转座酶/整合酶蛋白。这些酸性氨基酸构成了部分的活性位点以及协调反应活性所需的二价金属离子（如 Mg^{2+} 或 Mn^{2+}）（同 DNA 聚合酶，第 8 章）。这类蛋白质的一个显著特点是它们用这一个位点既催化 DNA 断裂又催化 DNA 单链转移，而不是用两个活性位点来分别催化这两个化学反应。

除了结构高度保守的催化结构域外，其他区域在这个蛋白质家族中是不保守的。这些非保守区域编码位点特异性的 DNA 结合域以及参与蛋白质相互作用的区域，这是组装特异性针对各个重组因子的蛋白质- DNA 复合体时所必需的。这样，这些独特的区域确保了转座酶或整合酶只催化编码自己的转座因子或是与其高度相关因子的重组。

转座酶和整合酶只有结合在 DNA 上形成突触复合体（或叫转座体）时才有活性（如前所述）。Tn5 转座酶结合到一对转座子末端 DNA 片段时的晶体结构让我们清楚地了解到这一点（图 11-25）。与这些 DNA 片段（即一个转座子的末端）其中一个相结合的重组酶识别序列上的转座酶亚基，为转座子另一端的 DNA 断裂和 DNA 单链转移反应提供了催化结构域。正是由于这个亚基的结构，只有在两个亚基和一对 DNA 末端一起构成复合体时，转座酶才被放置到正确的重组位置。

聚腺苷酸（poly-A）反转录转座子通过一种“反向剪切”机制进行移位

聚腺苷酸反转录转座子，如人的 LINE 因子的移位需要一个 RNA 中间体，但其转座机制与类病毒转座因子有所不同，它的机制叫做靶位点引导反转录（target site primed reverse transcription）（图 11-26）。首先是细胞内 RNA 聚合酶对整合因子的转录（图 11-26a）。尽管启动子在 5′UTR 内，这时它能指导起始于转座因子序列的第一个核苷酸的 RNA 合成。

新合成的 RNA 被转运到细胞质中翻译成 ORF1 和 ORF2 两个蛋白质（见上面的讨论）。这些蛋白质与编码它们的 RNA 保持结合（图 11-26b）。转座因子以这种方式促进了自身的转座，而不为其他竞争因子贡献蛋白质。

这个蛋白质-RNA 复合体重新进入核内并与细胞 DNA 结合（图 11-26c）。已知 ORF2 蛋白同时具备核酸内切酶和反转录酶的活性，这个核酸内切酶通过在染色体 DNA 上产生一个缺口来启始整合反应（图 11-26d）。富含胸腺嘧啶（T）的序列是具断裂起向的位点。在断裂位点出现的多个 T 使得 DNA 与 RNA 因子的聚腺苷酸尾巴形成碱基配对，然后切口反应产生的 DNA 3′-OH 末端作为引物反转录 RNA 转座因子（图 11-26e）。另外，ORF2 蛋白也催化这一 DNA 合成反应。后续的转座步骤尽管了解地还不太清楚，但是应包括 cDNA 第二单链的合成、插入位点 DNA 缺口的修复和 DNA 单链的接合。

通过大规模基因组测序发现很多聚腺苷酸反转录转座子都被截断了，其中大部分都丢失了 5′端的区域，而不具备编码因子基因的完整拷贝或完整的启动子，因而这些被截断的转座因子失去了转座能力。

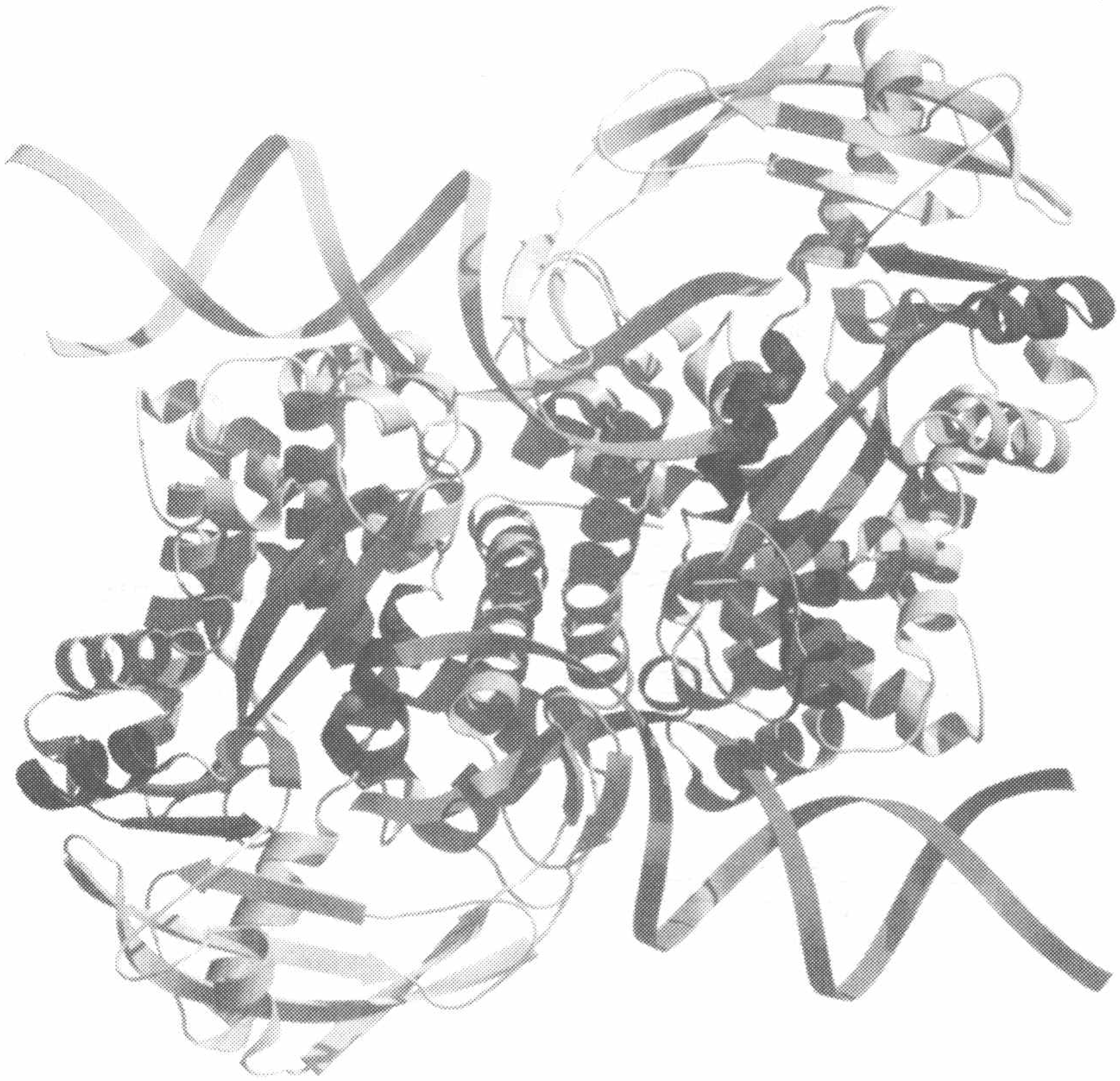


图 11-25 Tn5 结合 DNA 底物后的共结晶体。复合体中包含一个转座酶的二聚体。在图 11-24 中催化结构域以彩色表示，绿色小球表示与蛋白质活性位点结合的二价金属离子。注意当重组酶亚基的 DNA 结合域连接到转座子一端时，其重组催化结构域也贡献给转座子的另一端。DNA 用蓝色和粉色表示。(Davies D. R et al. 2000. Science 289：77-85.）图形通过 BobScript、MolScript 和 Raster 3D 软件制作，Leemor Joshua-Tor 完成了额外的 DNA 建模。

转座子及其调控的实例

转座子成功地侵入和繁殖于所有生物体的基因组中。很明显，它们是非常健壮的生物实体。这些成功可以部分归功于转座过程的调控，正是这种调控使得转座子与宿主细胞之间建立了一种和谐的共生关系。这种共生关系对转座子来说是至关重要的，因为没有了宿主，转座子就无法存在。另一方面，就像刚才讲的，转座子也会在宿主细胞中造成严重破坏，造成插入突变、改变基因表达、形成大规模基因重排。这些干扰破坏在植物中尤其明显，正是这一点导致在玉米中发现了转座子（框 11-4，玉米转座因子及转座子的发现）。

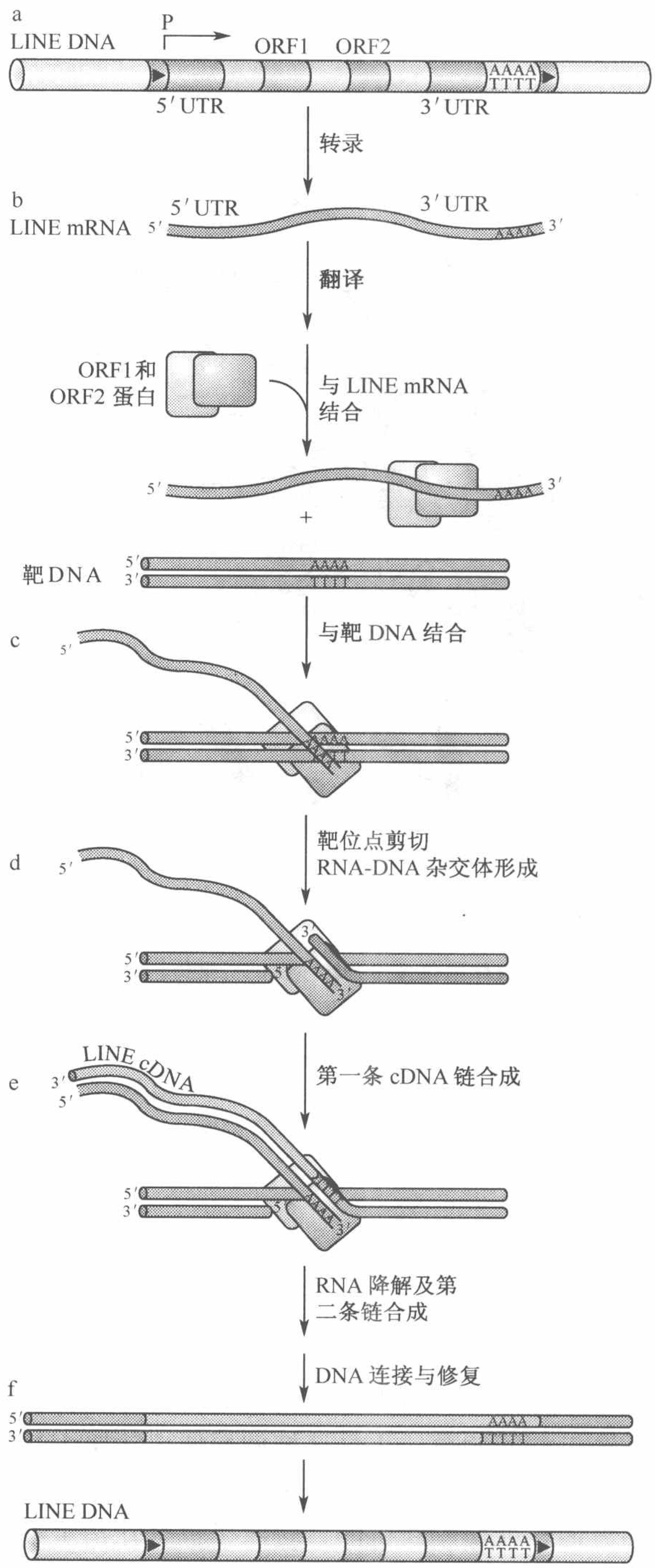
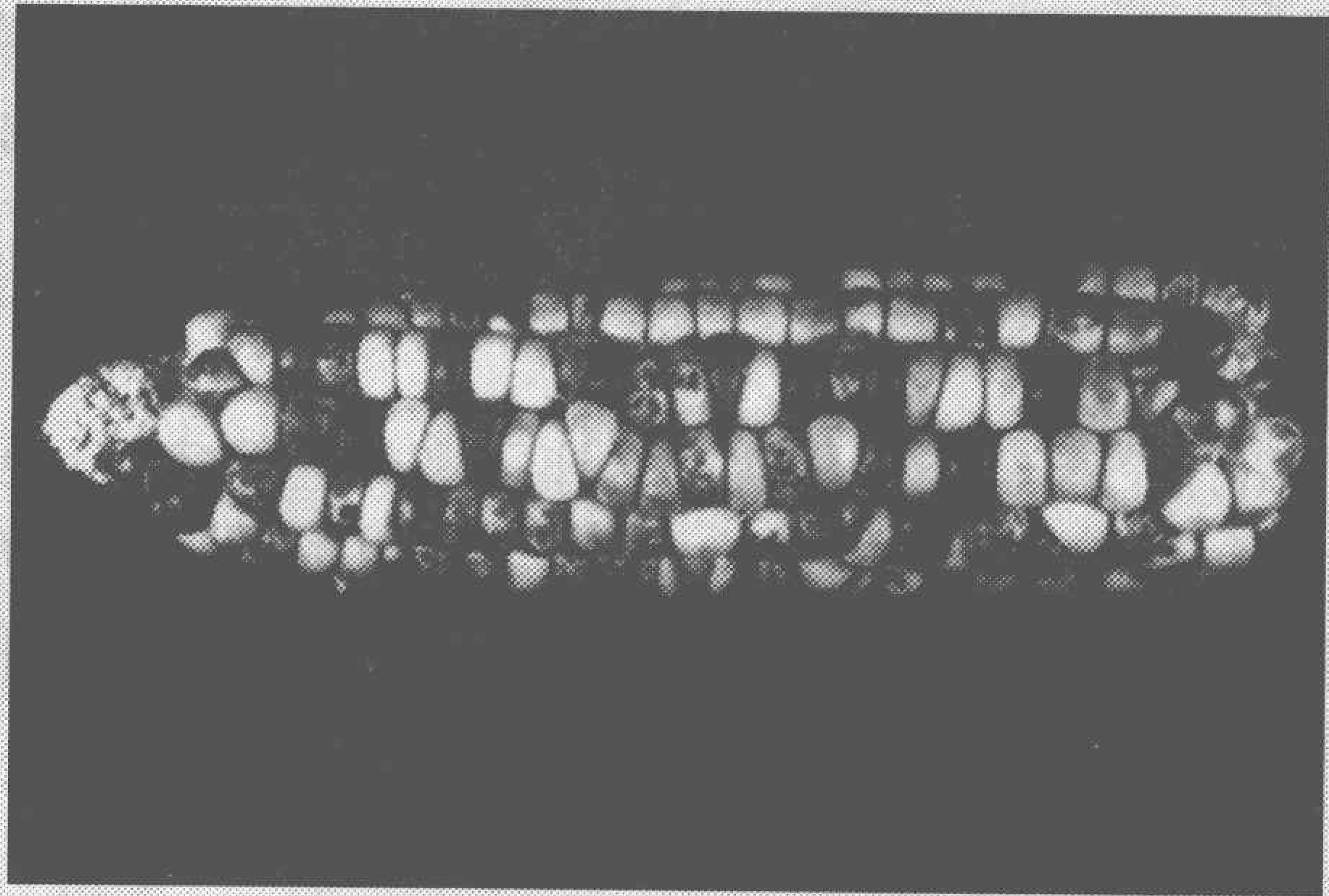


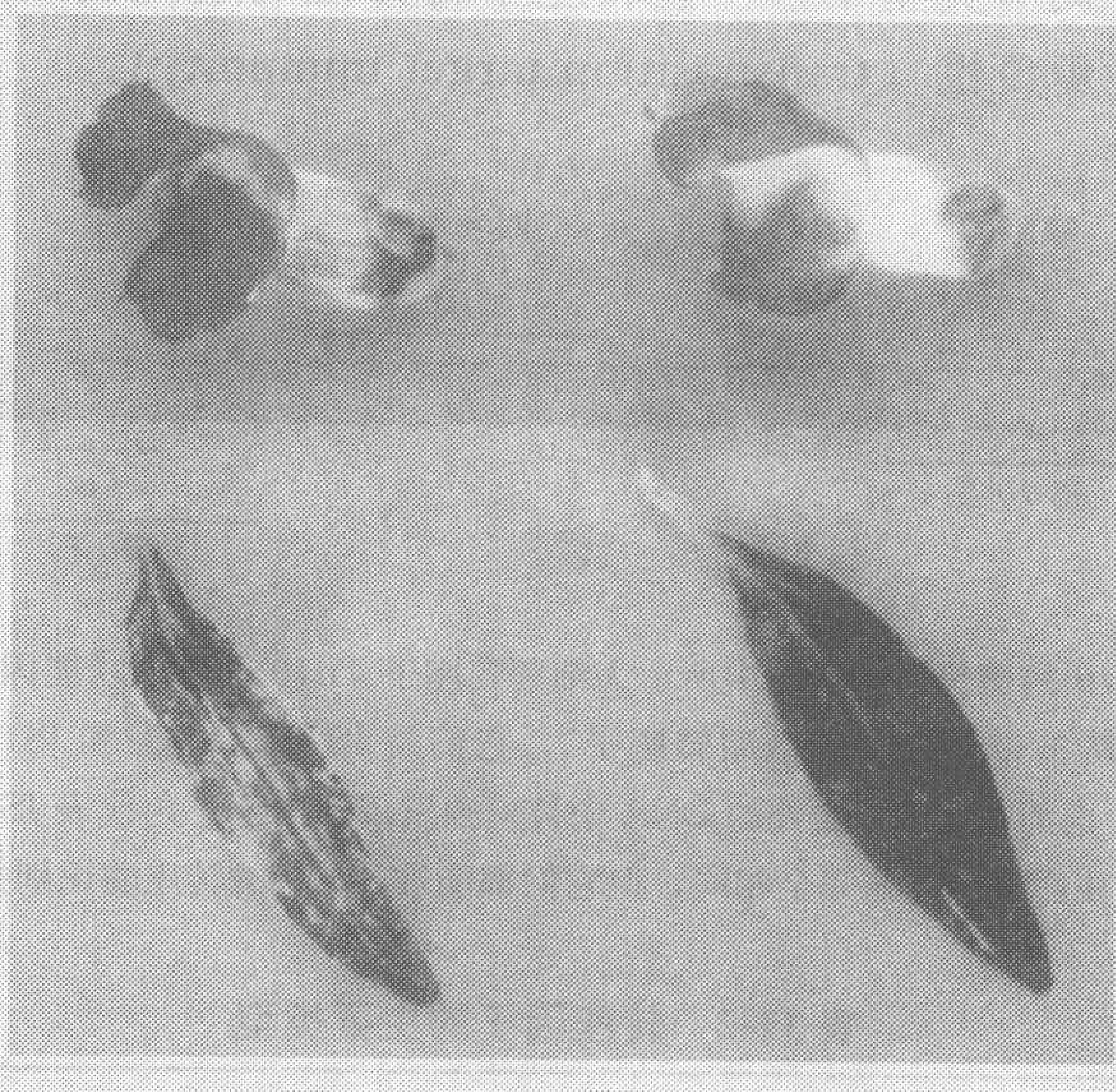
图 11-26　聚腺苷酸反转录转座子的转座通过靶位点引导的反转录过程完成。图示概述了 LINE 因子的移动模型。a. 细胞内的 RNA 聚合酶起始了一个整合的 LINE 因子的转录。b. 转录得到的 mRNA 翻译后生成两个 ORF 产物，并结合到其 mRNA的 3′端。c. 这个蛋白质-mRNA 复合体结合到富含胸腺嘧啶的靶 DNA 位点。d. 复合体上的蛋白质起动了靶 DNA 的断裂，在 DNA 末端留下了 3′-OH，并形成 RNA∶DNA 的杂合子。e. 靶 DNA 的 3′-OH 端作为引物反转录 RNA 因子得到 cDNA（第一条单链的合成）。f. 最后的转座反应步骤包括：第二条 cDNA 单链的合成、DNA 的连接以及断点修复，最终生成了一个新插入的 LINE 因子。

框 11-4 玉米转座因子及转座子的发现

植物基因组中含有大量的转座子。而且，转座因子改变基因表达的功能逐渐通过植物体颜色的变化被观察到（框 11-4 图 1）。因此，在植物中首先发现转座子和许多它们的沉默特征就不奇怪了。



框 11-4 图 1A 玉米棒上因转座而出现彩色色斑的例子。（由 Barbara McClintock 拍摄，版权归 Cold Spring Harbor Laboratory Archives.）



框 11-4 图 1B Tam3 转座引起金鱼草花出现彩色色斑的例子。白色斑块的大小与转座频率有关。（征得许可后重新印刷，Chatterjee M. 和 Martin C. 1997. Plant J. 11：759-771；Fig. 2a. © Blackwell Publishing）。

20 世纪 40 年代后期，Barbara McClintock 在玉米中发现了“控制因子”，实际上是转座因子断裂染色体的能力引起了 McClintock 的注意。她发现一些植株的染色体频繁断裂，并将造成这些染色体断裂的遗传因子命名为 *Ds*（dissociator；解离子）。令人

惊奇的是，她观察到这些染色体断裂的“热点”位点在不同的植株上位置不同，甚至一个植株和它自己后代之间的位点都不同。这一现象首次揭示了遗传因子可以在染色体上移动，即“转座”。

Ds 实际上是一个非自主性的 DNA 转座子，它通过剪切-粘贴转座实现移动。McClintock还发现 *Ds* 的移位需要 *Ac*（activator；激活子）因子存在于同一细胞，并提供转座酶蛋白。*Ac* 因子现在已确定属于一个叫 *hAT* 的 DNA 转座子大家族，其名称是苍蝇的 *hobo* 因子、玉米的 *Ac* 因子和金鱼草的 *Tam* 因子的首字母缩写。家蝇中的 Hermes 因子也属于这个家族并已用在机理分析中。

下面我们简要介绍几个已经充分了解的转座子和转座子家族（更多转座子及其重要特性被总结在表 11-2 中）。每一个小节对一个特定的因子进行简要介绍，并介绍一个关于该因子调控的例子，这种调控作用对该转座子来说非常重要。我们将会看到，有两种调控类型会反复出现：

- 细胞中转座子对自身拷贝数的控制。转座因子通过调节拷贝数，限制自身给宿主细胞带来的有害影响。
- 转座子对靶位点选择的控制。已经发现了两类常见的靶位点调控方式：其一，一些因子倾向于插入到对宿主细胞不造成危害的染色体区域，这些区域被称为转座子的避风港；在第二种调节类型中，一些转座子特异性的避免插入到自身序列中，这种现象叫做转座目标免疫性（transposition target immunity）。

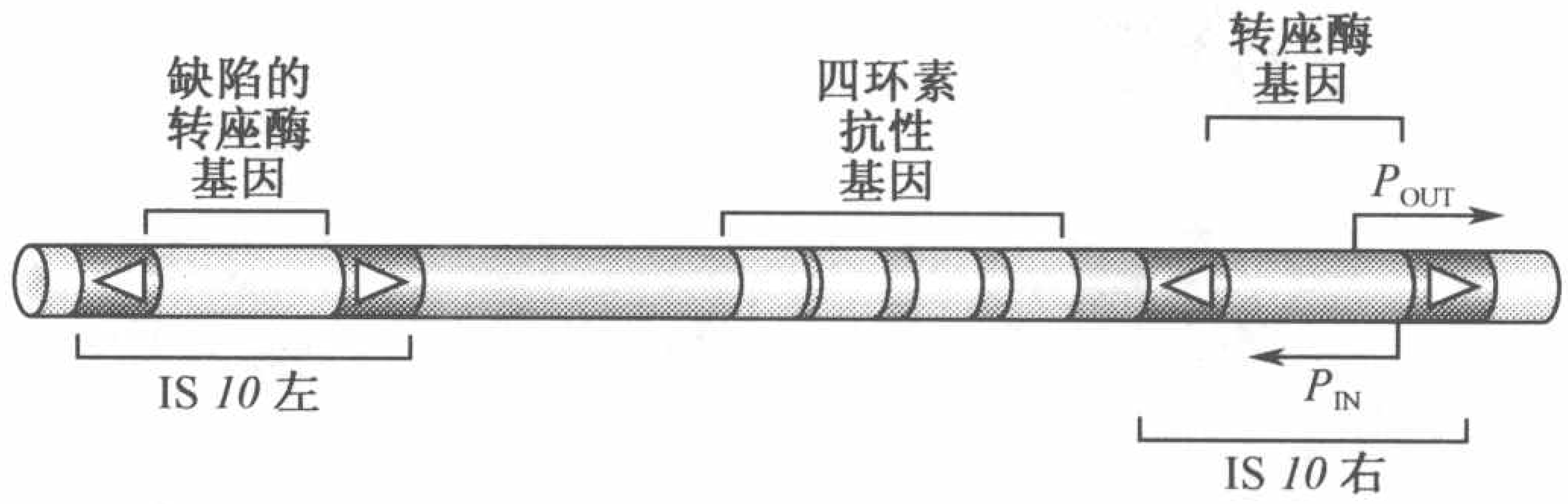


图 11-27　细菌转座子 Tn*10* 的遗传结构。图中所示为细菌转座子 Tn*10* 的各个功能因子。像很多细菌转座子一样，Tn*10* 实际上在其两端携带了两个“微型转座子”，它们是 IS*10*L（左）和 IS*10*R（右）。这两种类型的 IS*10* 因子都能自行转座，也都在 Tn*10* 之外的 DNA 中被单独发现。白色三角代表 IS 因子和 Tn*10* 末端的反向重复序列。虽然 4 个拷贝在序列上并非完全相同，但它们都能被 Tn*10* 的转座酶识别并用作重组位点。

表 11-2　转座因子的主要类型

类型	结构特点	移位机制	例子
DNA 介导的转座			
细菌的复制性转座子	末端反向重复序列包围着中间的抗生素抗性基因和转座酶基因	每次插入到一个新的靶位点都伴随着 DNA 因子的复制	Tn*3*、γδ、Mu 噬菌体

续表

类型	结构特点	移位机制	例子
细菌的剪切-粘贴转座子	末端反向重复序列包围着中间的抗生素抗性基因和转座酶基因	DNA因子从原位点被切除后，插入到一个新位点	Tn*5*、Tn*10*、Tn*7*、IS*911*、Tn*917*
真核细胞转座子	反向重复序列包围着带有内含子的编码区	DNA因子从原位点被切除后，插入到一个新位点	P因子（果蝇）、*hAT*家族因子、Tc*1*/*Mariner*因子
RNA介导的转座			
类病毒反转录转座子	250～600bp的正向末端重复序列（LTR）包围的反转录酶、整合酶和类反转录病毒的Gag蛋白的基因	RNA聚合酶Ⅱ从左边LTR的启动子开始转录出RNA，然后再反转录并插入到靶位点	Ty因子（酵母）、*Copia*因子（果蝇）
聚腺苷酸反转录转座子	3′端富含A-T的序列和5′端UTR序列包围着编码RNA结合蛋白和反转录酶的基因	从其内部的启动子开始转录生成RNA；核酸内切酶切断DNA起始靶位点引导的反转录过程	F和G因子（果蝇）、LINE和SINE（哺乳动物）、*Alu*序列（人）

IS4家族的紧密型转座因子通过多种机制实现对拷贝数的控制

细菌转座子Tn*10*是IS*4*家族中一个被充分研究的代表，这个家族还包括Tn*5*。Tn*10*是一个9kb长的紧密型转座因子，它编码一个自身转座酶所需的基因和几个使其对四环素具有抗性的基因（图11-27）。

Tn*10*转座子采用剪切-粘贴机制（如前述），通过DNA发卡结构断裂非转移单链（图11-18和图11-21）。Tn*10*序列上还有一个IHF结合位点，IHF帮助重组所需的转座体复合物正确地组装，IHF在λ细菌噬菌体整合过程中也起到这个作用（如前述）。

Tn*10*由3个功能模块组成，这个结构有一定的普遍性，具备这种结构的因子被称为复合转座子（composite transposon）。最外面的两个模块叫IS*10*L（左边）和IS*10*R（右边），实际上是微型转座子。“IS”代表的意思是插入序列（insertion sequence）。IS*10*R是编码转座酶的基因，转座酶能够识别IS*10*R、IS*10*L和Tn*10*的末端反向重复序列。IS*10*L的序列虽然与IS*10*R非常相似，但它不能编码具有功能的转座酶。所以，IS*10*R和Tn*10*是自主转座子，而IS*10*L则是非自主转座子。如预期的一样，由于自身可移动，两种类型的IS*10*转座因子在基因组不与Tn*10*相连。

Tn*10*在某种指定细胞中通过限制转座发生的频率来限制其拷贝数。它使用一种反义RNA（antisense RNA）的机制来限制转座酶基因的表达（图11-29）（第17章和第18章关于反义RNA调节的讨论）。靠近IS*10*R末端的是两个启动子，它们通过宿主RNA聚合酶指导RNA合成。指导RNA向内合成的启动子（称作P_{IN}）负责转座酶基因的转录，而指导向外转录的启动子（称作P_{OUT}）如下所述通过生成反义RNA来调节转座酶的表达。从P_{IN}和P_{OUT}转录的RNA有一个重叠区（大约36bp），在这个重叠（互补）区可以通过氢键配对阻止核糖体结合到P_{IN}转录的RNA，也就阻止了转座酶蛋白的合成。

通过这个机制，细胞内的Tn*10*拷贝数越多，反义RNA也就越多，反过来又限制

了转座酶基因的表达（图 11-28，更多的细节参照图示），因而，这个细胞株中的转座频率就会很低。相反，如果细胞中仅有一个 Tn*10* 拷贝，反义 RNA 的水平就会很低，转座蛋白就会高效地合成，转座发生的频率就较高。

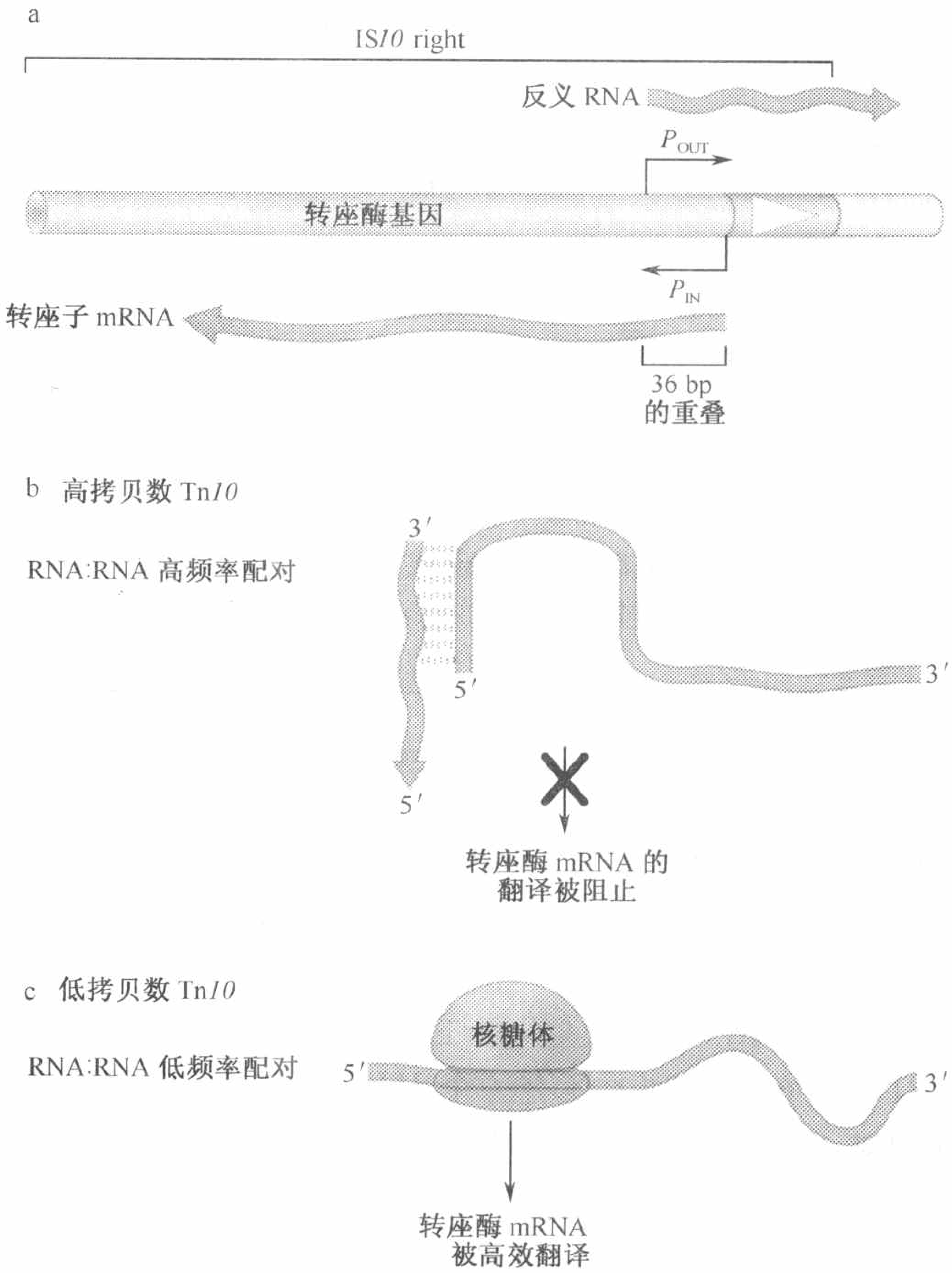


图 11-28 Tn*10* 表达的反义调节。a. 重叠的启动子区域如图所示，向左的启动子（P_{IN}）启动转座酶基因的表达，向右的启动子（P_{OUT}）开始于距 P_{IN}启动子左端 36bp 处，它启动反义 RNA 的表达。这两个转录本前 36bp 是互补的。注意在细胞中起始于 P_{OUT}的反义转录本较起始于 P_{IN}的转录本有较长的寿命。b. 细胞中 Tn*10* 的拷贝数多时，这两个 RNA 的配对就频繁发生，阻止了转座酶 mRNA 的翻译（因而最终减少了转座子的拷贝数）。c. 当细胞中转座子拷贝数很低时，两种 RNA 转录本配对很少，转座酶 mRNA 能够高效翻译，细胞中转座子的拷贝数增加。

Tn*10* 的转座伴随着细胞 DNA 的复制

Tn*10* 的转座也伴随着细胞 DNA 的复制。前面讲过细菌如 *E. coli*（为 Tn*10* 的常见宿主）往往会在 GATC 位点甲基化（第 8 章，框 8-7）。甲基化发生在 DNA 复制后，所以在复制叉经过和被甲基化酶识别之前，这些 GATC 位点有几分钟处于半甲基化状态。

就是在 Tn*10* DNA 半甲基化的这个短短的期间内，转座最可能发生。在甲基化状态联合转录是由于转座子序列上有两个关键的 GATC 位点，一个位于转座酶基因的启动子内，另一个位于结合转座酶的末端反向重复序列内。RNA 聚合酶和转座酶对半甲

基化序列的结合比它们对完全甲基化序列的结合要强，所以当DNA处于半甲基化状态时，转座酶基因最高效表达，转座酶蛋白最有效地结合DNA。结果，在细胞周期中DNA复制后的短暂期间内，Tn*10*转座的发生频率最高（图11-29）。

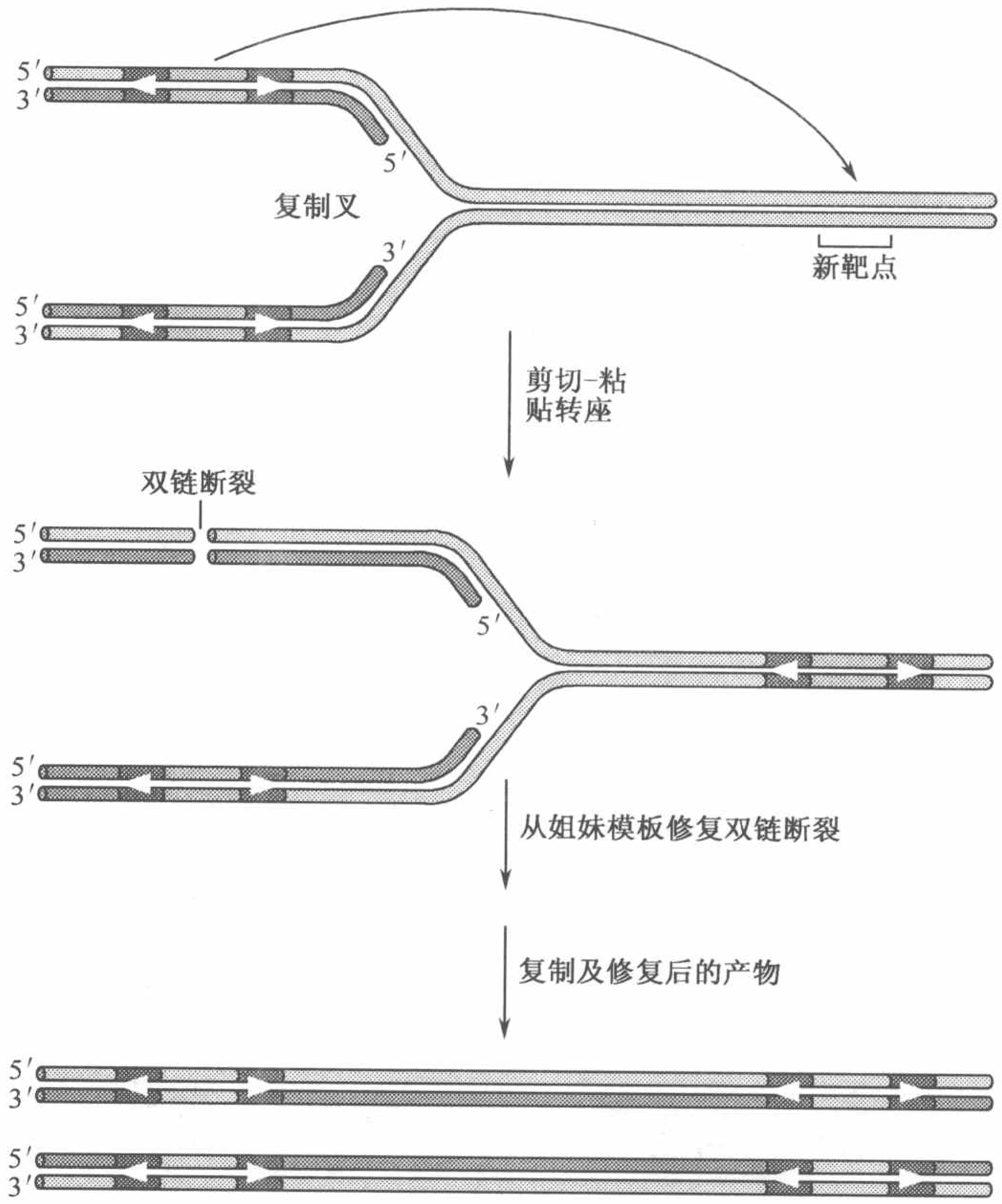


图11-29 复制叉经过后的Tn*10*转座。转座过程在DNA复制后有半甲基化DNA存在时被激活（甲基化位点没有显示）。在转座过程中，在染色体上转座因子被切除的位置形成了一个双链DNA断裂，这个断裂被DSB修复途径修复（第10章），这一过程在切除位点产生了一个Tn*10*的拷贝。利用这个机制，转座过程实际上表现出了“复制性”，尽管转座过程的确是通过剪切-粘贴（非复制性）途径完成的。

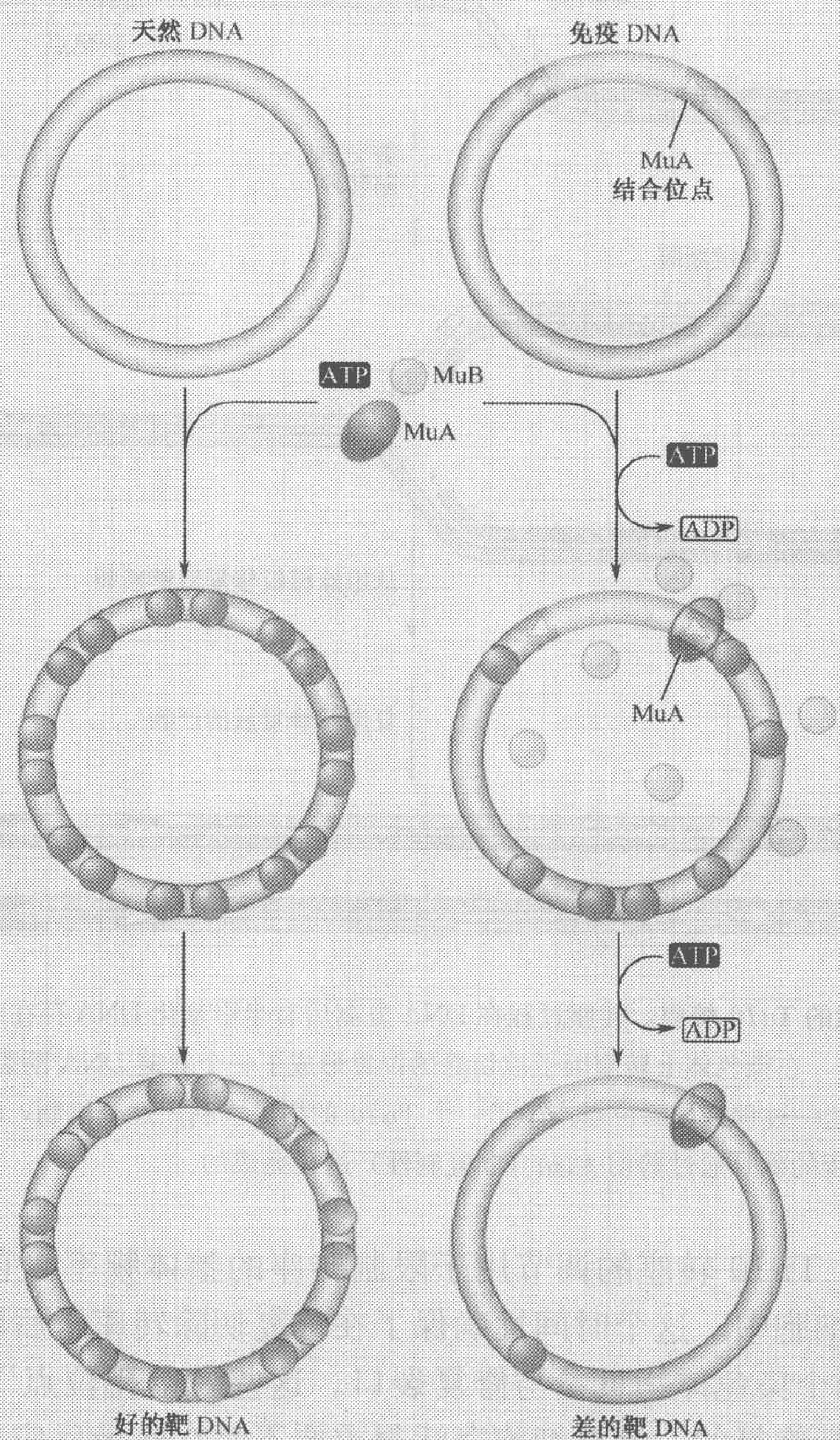
DNA甲基化对Tn*10*转座的调节用于限制转座的整体频率。它也特别限制了转座发生在活跃的分裂细胞中。这个时间性确保了在修复切除转座子后旧位点上留下的双链DNA断裂时，有两个染色体拷贝参与修复裂口。这些“空靶位点”是通过同源重组双链断裂修复途径进行修复的，该对细胞存活很重要不是在转录反应中该染色体区域有两份拷贝（第10章）。

框11-5 转座的靶免疫机制

转座目标免疫机制的核心是MuA转座酶和MuB ATP酶间的相互作用。正是MuA和MuB的相互作用阻止了MuB结合到与MuA结合的DNA附近。这个相互作用对它们的影响是：

- MuA 阻止了 MuB 结合到其附近的 DNA 位点。这一过程需要 ATP 的水解。
- MuB 帮助 MuA 找到转座的靶位点。

要了解单个的蛋白质-蛋白质和蛋白质-DNA 相互作用如何一同产生目标免疫，我们来看看转座到两个候选 DNA 片段时的情况：一个代表任意的 DNA 片段，而另一个片段带有一个插入的 Mu 因子的拷贝（框 11-5 图 1）。我们把第一个片段叫天然区域，把第二个片段叫免疫区域。



框 11-5 图 1 MuA 和 MuB 在 DNA 上的相互作用导致了靶位点免疫的发生。MuA 结合位点位于转座子末端的末端反向重复序列中（深绿色表示），为了说明清楚，图中所示的 MuA 仅结合了两个重复区域中的一个。每当 MuB 水解 ATP，它就从 DNA 脱了下来（结合 ATP 的 MuB 以更深的绿色显示）。MuA 和 MuB 的相互作用能促进这个水解反应，尽管图中所示只有两个 MuB 分子与 MuA 作用。实际上 MuA 总是趋向于接触到所有的紧密连接在它的 DNA 结合位点上的 MuB 分子。一个单独的 MuA 结合的末端反向重复序列可以使 5～15bp 的 DNA 序列获得“免疫性”。

那么当Mu因子准备转座时这两个DNA区域时会发生什么呢？首先考虑天然区域，结合ATP形成复合体（MuB-ATP）的MuB蛋白以非特异性的方式结合此DNA上，与此同时，MuA转座酶在MuDNA上组装转座体，然后转座体内的MuA蛋白与天然区域上的MuB-DNA复合体发生蛋白质-蛋白质相互作用。这种相互作用的结果是MuB将这个序列作为靶位点传递给MuA。

与此相反，当MuA和MuB都结合到免疫区域的DNA上时，MuA与免疫区域上Mu噬菌体基因组的特异性位点发生相互作用。MuB-ATP的结合仍采用非特异性的方式。然而当MuA和MuB都结合到这一区域时，它们将发生相互作用，结果是MuA通过MuB刺激ATP水解，使得MuB与DNA分离，因此，MuB就不能在免疫DNA片段上聚集。通过这个方法，Mu转座蛋白用ATP中保存的能量来保护其基因组免于成为转座的目标。从这个机制中我们还可以预测出，只要DNA上有一个单独的MuA结合位点就足以获得目标免疫性。

Mu噬菌体是一种异常活跃的转座子

和λ噬菌体一样，Mu噬菌体也是一种溶原性细菌噬菌体（第22章）。Mu也是一种较大的DNA转座子，它侵染宿主细胞时通过转座将自己的DNA插入到宿主的染色体上，所用的方式与反转录病毒类似（如前述）。Mu噬菌体在裂解性生长时也通过多轮的复制性转座扩增自身DNA。在裂解周期中，Mu噬菌体每小时能完成大约100轮的转座，它是目前已知的最高效的转座子。而且即使在静止溶原态，与传统的转座子（如Tn*10*）相比，它的转座频率仍然是相当高的。它的名字Mu是突变子（mutator）的缩写，这源于其疯狂转座的能力，因为当细胞内插入一个拷贝的Mu噬菌体DNA后，噬菌体DNA插入细胞基因中而使细胞频繁地增加新的突变。

Mu噬菌体基因组大小约为40kb，带有35个以上的基因，其中仅有两个基因编码的蛋白质专注地参与转座，它们是基因*A*和基因*B*，分别编码MuA蛋白和MuB蛋白。MuA是一个转座酶，属于前面讲过的DDE蛋白超家族的成员；MuB蛋白是一个ATP酶，它促进MuA的活性，并控制对DNA靶位点的选择（图11-30）。这个过程将在下一节介绍。

Mu噬菌体通过靶位点免疫来避免其转座到自身DNA上

像许多转座子一样，Mu噬菌体对靶位点也很少显示出偏好性。结果，“好”的靶点常常出现在DNA上，Mu基因组自身的DNA也被频繁地选为靶位点。如果序列的选择如此随机，转座子插入到噬菌体自身DNA上后，很容易造成噬菌体基因的严重紊乱，那么Mu噬菌体是如何避免这一情况呢？

由于Mu转座子受到转座目标免疫（transposition target immunity）（框11-5）作用的调控，使得这一问题得以解决。在转座时，这个作用致使Mu因子周围的DNA位点以及包括转座子自身的位点，不易成为转座的靶位点。

已经发现了很多不同的可转座因子都具有转座目标免疫性，而且可以远距离发生作

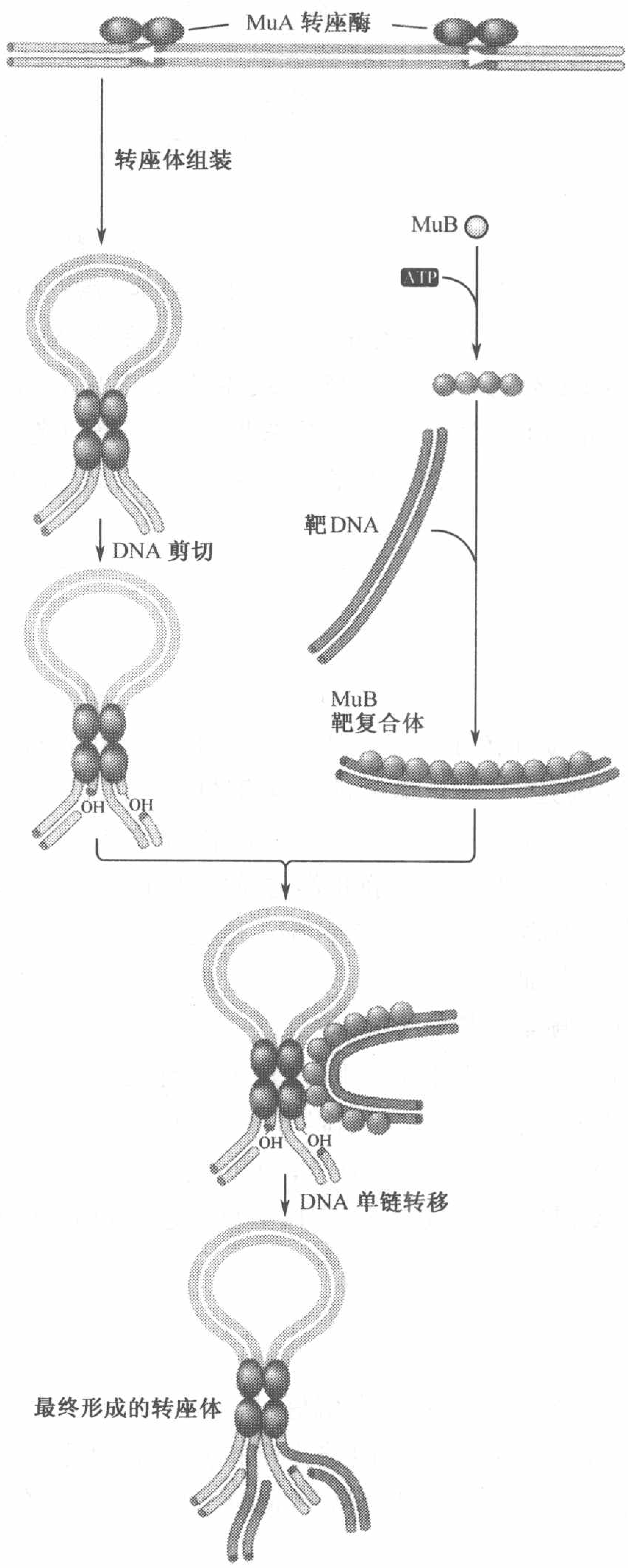


图 11-30 Mu 转座早期步骤的概述。MuA 转座酶的 4 个亚基结合到 Mu DNA 的末端，MuB 蛋白结合 ATP 再结合到任意的 DNA 序列上。然后，MuA 和 MuB 蛋白之间的相互作用将 MuA DNA-转座体复合体带到一个新的 DNA 靶位点。MuB 在最后一个图中没有显示，这是因为在 DNA 单链转移后便不再需要它，它可能离开了复合体。

用。例如，一个已经存在的 Mu 转座子可以对附近约 15kb 的范围实现 Mu 转座子对新插入的免疫。还有一些转座因子（如 Tn*3* 和 Tn*7*）的目标免疫范围超过了 100kb。目标免疫使得一个转座因子免于被自身或同类型因子的另一个拷贝插入。而且，这种对靶

DNA 选择的调节推动了转座因子移动到远离自身的新位置，这个特点或许有益于转座子整体的生存和繁衍。

Tc*1*/*mariner* 因子是真核细胞中极其成功的 DNA 转座因子

已经发现的 Tc*1*/*mariner* 转座因子家族成员在非脊椎动物和脊椎动物中都普遍存在，这个家族的转座因子是真核细胞中最常见的 DNA 转座子。尽管这些因子有明显的相关性，但来自不同物种的成员有着不同的特性和命名。例如，来自线虫 *C. elegans* 的因子叫 Tc 因子，而从果蝇体中分离出的因子叫 *mariner*。

Tc*1*/*mariner* 因子是已知最简单的自主转座子之一。通常，它们长度大约1.5～2.5kb，仅包含一对反向末端重复序列（转座酶结合位点）和一个编码转座酶蛋白的基因，该转座酶属于 DDE 蛋白超家族（如前述）。与很多转座子不同的是，它的转座过程不需要辅助蛋白的参与，尽管在最后的重组步骤中需要细胞的 DNA 修复蛋白。或许如此简单的结构和机制正是它们在如此广泛的宿主物种中获得巨大成功的原因。

Tc*1*/*mariner* 因子的转移是通过剪切-粘贴的转座机制实现的（图 11-19）。转座子 DNA 从原宿主 DNA 两端的切除是通过单链的成对断裂实现的，断裂位点间隔两个碱基。这些因子非常容易插入到有 5′TA · 序列（很明显常见的序列）的地方。

转座子被切除后在宿主染色体上留下的“空”位点怎么办呢？在 Tc*1*/*mariner* 转座中，对一些曾经携带过转座子的位点分析表明，有时这些断开的 DNA 末端被修补（通过 DNA 的修复合成）后就直接连接起来（第 9 章，关于非同源末端连接的讨论）。这些修复反应会在原先的位点插入几个额外的碱基对，这些短小的插入 DNA 被称为“脚印”，因为它们是转座子“途经”这里后在基因组上留下的踪迹 。

与很多转座子不同，Tc*1*/*mariner* 因子的转座并未得到很好的调控。或许就是由于缺乏调控，在基因组序列上发现了很多“死”的转座子，即无法转座。例如，很多转座酶基因发生突变造成转座子失活。研究人员用活性因子和无活性因子上的很多序列人为构造了一个极度活跃的 Tc*1*/*mariner* 因子，称为“睡美人”。与分离出的天然因子相比，它能以很高的频率发生转座。“睡美人”很有希望成为一个用于在真核物种细胞中诱发突变和插入 DNA 的有效工具。而且这个重构实验还说明，天然的 Tc*1*/*mariner* 因子由于其转座酶蛋白没有发挥最佳作用而自然降低了转座的频率。

酵母 Ty 因子转座到基因组的安全港

Ty 因子（酵母中的转座子）是酵母中重要的转座子，属于类病毒反转录转座子。事实上，它们与反转录病毒的相似性已经超越了转座机制。Ty RNA 在细胞中被发现组装成类病毒的颗粒（图 11-31），所以这些因子好像是病毒一样，只是不能离开一个细胞去感染另一个细胞。已经有很多类型的 Ty 因子被充分研究。例如，酿酒酵母中有 Ty1、Ty3、Ty4 和 Ty5 因子（尽管 Ty5 在这类酵母中都是失活的）。每一类 Ty 因子都能催化自身的移位，但不能移动其他类的因子。

Ty 因子偏向于整合到染色体上的特定区域（图 11-32）。例如，Ty1 因子几乎总是

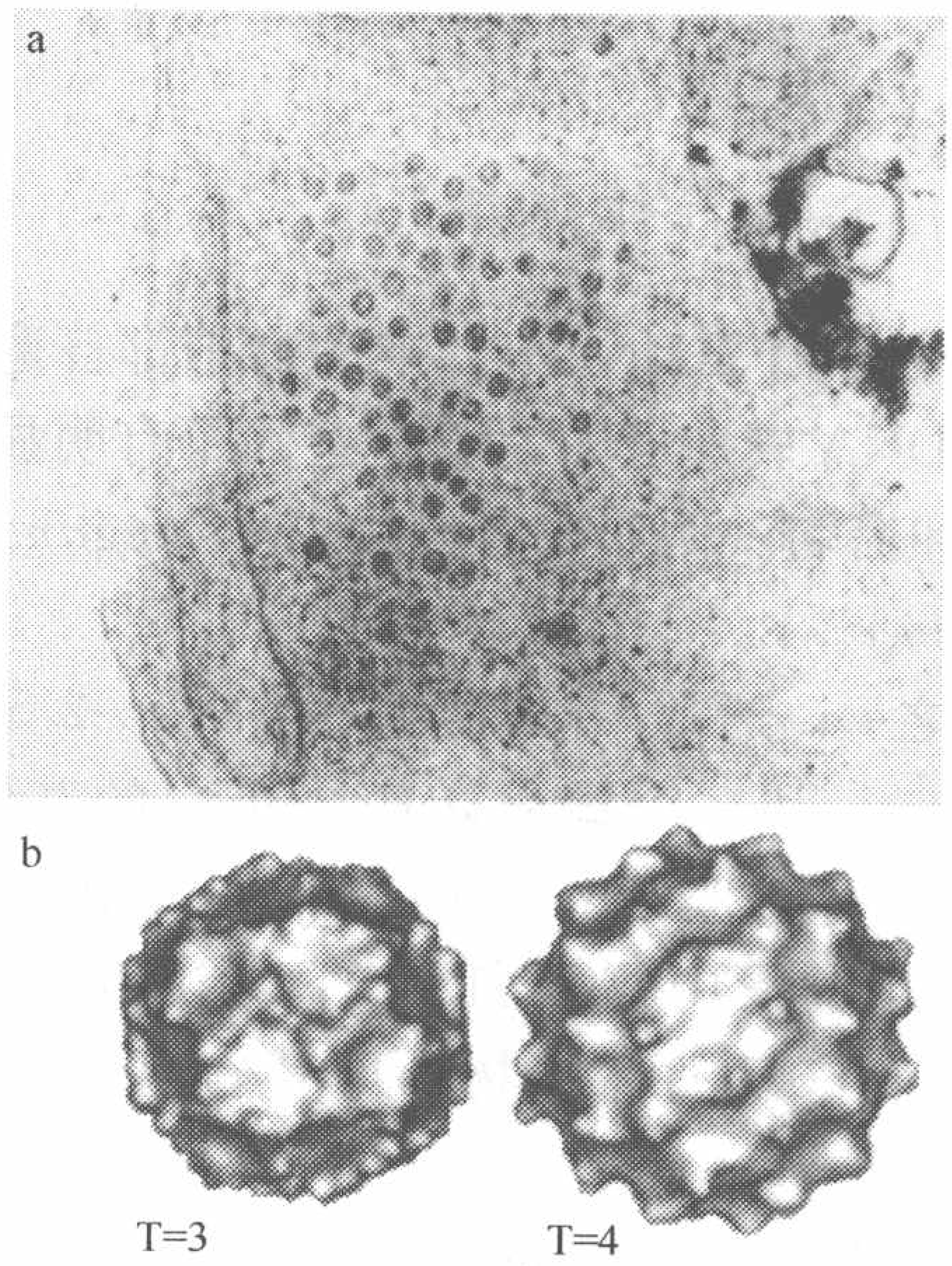


图 11-31 酵母的 Ty 因子被组装到病毒颗粒中。 a. 过表达的Ty1类病毒颗粒的酿酒酵母细胞的电子显微照片。颗粒呈卵形、电子密集的结构。b. 低温电子显微扫描显示的Ty1病毒颗粒的三维重构结构。Ty1因子所携带的被切断的Gag蛋白与病毒颗粒的三聚体构成了钉状外壳。（来源：Craig N. et al. 2002. Mobile DNA Ⅱ，a. ASM Press.）

插入到被宿主RNA聚合酶Ⅲ转录的起始位点上游约200bp的区域内（第12章）。RNA聚合酶Ⅲ特异性地转录tRNA基因，所以大多数Ty1因子就插入到这些基因附近。Ty3因子的插入也和RNA聚合酶Ⅲ的启动子紧密相关。这一情况下，插入位点就精确定位在转录起始点位置（±2bp）。与此不同的是，Ty5因子趋向于插入到基因组的静止区，也就是转录静止区域，Ty5插入的静止区包括端粒和交配型基因座的静止拷贝（第10章）。在所有这些情形中，区域性靶位点的选择机制与一个特定的蛋白质-蛋白质复合体的形成有关，这个复合体由两个蛋白质组成，分别是与cDNA结合的复合体中的整合酶，以及与染色体位点结合的特定的宿主蛋白。例如，Ty5整合酶就与DNA沉默蛋白Sir4构成了一个特殊的复合体（第17章）。

为何Ty因子表现出这种靶位点的区域性偏好呢？有人提出，正是由于这种靶位点的选择性，使得大多数的插入远离基因组中编码蛋白的重要区域，让转座子能保留在宿主细胞中。使用目标性转座方式对酵母这种小而富含基因的基因组有更加重要的意义。

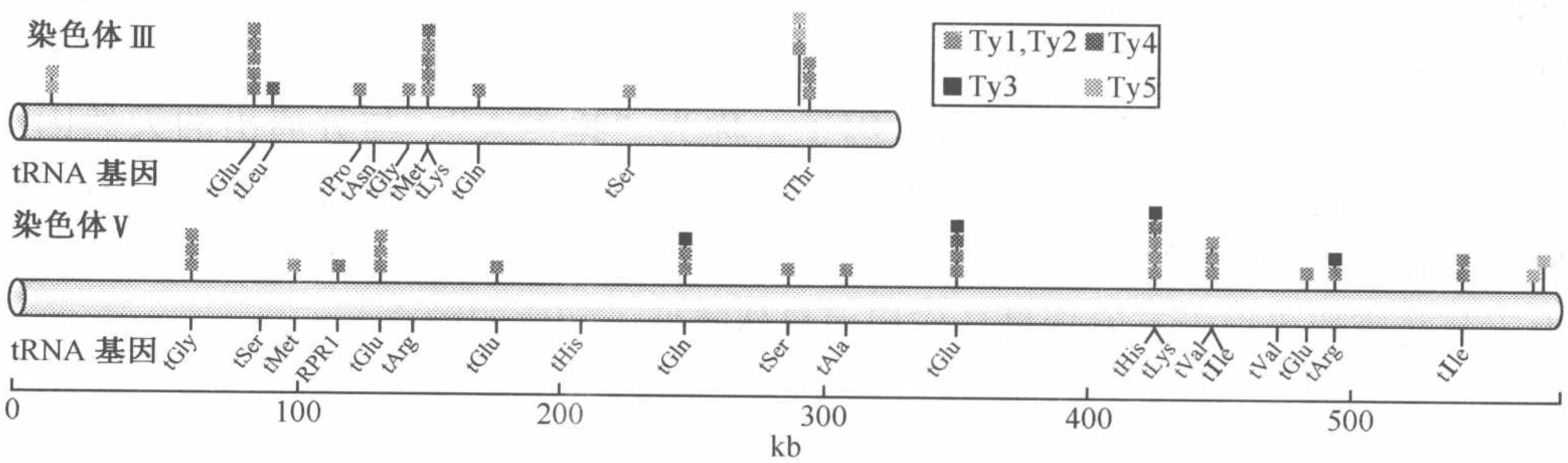


图 11-32 Ty 因子的整合位点簇。 每个彩色区块代表一个已知的转座插入位点。注意Ty1、Ty2、Ty3和Ty4都是插入到由细胞RNA聚合酶Ⅲ转录的tRNA基因附近，插入发生在真实基因的上游，因而不对基因的表达产生干扰。Ty1和Ty2是紧密相关的转座因子，所以归为一类；Ty5被发现位于染色体末端和交配型基因座（第10章）附近的"沉默"区域。（来源：Courtesy of Dan Voaytas.）

LINE 因子能够推动自身转座，还能转座细胞 RNA

自主性聚腺苷酸反转录转座子LINE在脊椎动物基因组中大量存在。事实上，人的

基因组中20%由LINE序列构成。这些因子起初是被当作重复序列家族发现的，它的名字来源于最初的定义：LINE是“散布在核内的长因子”（long interspersed nuclear element）的缩写。L1是人体内了解得最清楚的一个LINE因子。除了能推动自身的移位，LINE还能提供一些蛋白质，用于反转录和整合其他相关的重复序列，一种被称为“散布在核内的短因子”（short interspersed nuclear element，SINE）的非自主性聚腺苷酸反转录转座子。基因组序列显示这类因子也是大量出现，通常大小为100～400bp。在人的基因组中普遍存在的*Alu*序列就是SINE的一个例子。LINE和SINE因子典型结构的比较如图11-33。LINE和SINE的序列看起来像是简单的基因，可事实上，如对转座很重要的*cis*-acting因子，它仅包含一个指导该因子转录成RNA的启动子和一个聚腺苷酸序列。前面讲过，这些腺苷酸残基与靶位点的DNA配对，帮助形成反转录所需的末端引物（图11-33）。

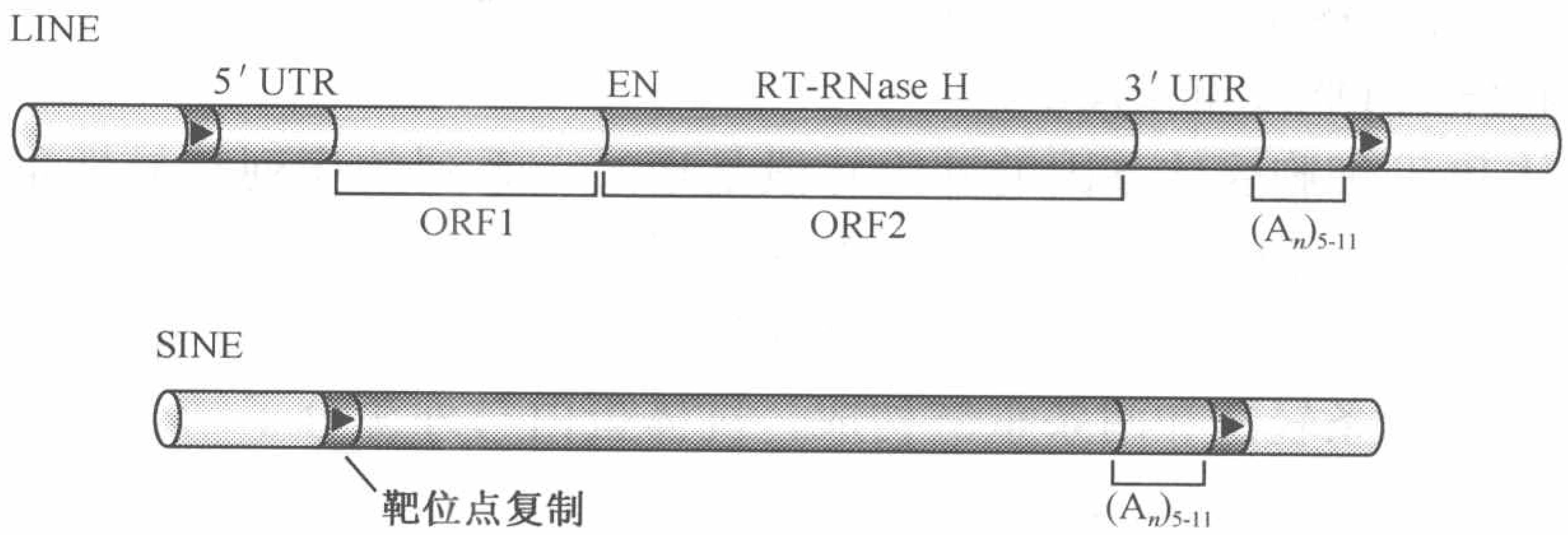


图11-33 一个典型的LINE和SINE的遗传结构。注意在因子右端的聚腺苷酸序列的长度是可变的。这是一个聚腺苷酸反转座子的典型特征这些因子两端的靶位点重复序列（蓝色箭头）的长度也是可变的。序列因子没有按比例显示。这两类因子都带有启动子序列，见图11-20和图11-28。（来源：Bushman F. 2002. Lateral DNA transfer，p. 251，Fig. 8.4. © Cold Spring Harbor Laboratory Press.）

这种转座所需的序列如此简单，给LINE因子提出一个问题：如何避免对细胞内mRNA的转座呢？我们知道，所有的基因都有启动子，它们被转录成mRNA后，在3′端都带有一个聚腺苷酸的序列（第12章），那么所有的mRNA都可能成为吸引转座酶的底物。实际上，基因组序列已经为我们提供了清楚的证据表明细胞RNA通过靶位点引导的反转录机制实现转座。

细胞中很多基因在基因组上都有一些高度同源的其他拷贝，这些拷贝丢失了启动子和内含子（存在于基因中的序列，在RNA剪切中被从mRNA中去掉，见第13章），通常在5′端被截短。这些序列被称为加工后的假基因（processed pseudogene），通常不能被细胞表达。这些假基因两端一般是靶DNA的短重复序列，这个结构恰好就是LINE促进的细胞mRNA转座的预期结果。

尽管细胞RNA会发生转座，但这是一个罕见的现象。避免这一过程的基本机制是编码LINE因子的蛋白质在翻译时迅速地结合到自己的RNA上（图11-23）。这样，它们在催化反转录和整合时，对编码自己的RNA表现出了很强的偏好性。

V(D)J 重组

我们看到转座涉及很多不同遗传因子的移动，然而细胞也在利用重组机制来实现对自身有帮助的功能。最好的例子就是 V (D) J 重组，它发生在脊椎动物免疫系统的细胞中。

脊椎动物免疫系统的工作是识别和防御生物体如病毒、细菌和病原细胞的入侵。脊椎动物用两类特殊的细胞（B 细胞和 T 细胞）专门识别这些入侵者。B 细胞制造循环于血液系统中的抗体（antibody）；而 T 细胞制造细胞表面结合的受体蛋白，叫 T 细胞受体（T cell receptor）。这两类蛋白质识别“外部”分子后便展开一系列旨在消灭入侵者的活动。为成功地完成它们的功能，抗体和 T 细胞受体必须能识别大量的多种分子。细胞用于产生多种多样的抗体和 T 细胞受体的基本机制所依靠的就是被称为 V (D) J 重组的一套特殊的 DNA 重排反应。

组成抗体和 T 细胞受体基因的基因片段是通过一系列的序列特异性重排装配而成的。要了解这个重组过程是如何产生所需的多样性，就需要了解一下抗体分子的结构（图 11-34）；T 细胞受体与抗体的分子结构相似。编码抗体分子的基因组区域如图 11-35 所示。抗体由两条轻链和两条重链构成，抗体蛋白与外源分子结合的部分叫抗原结合位点（antigen-binding site），这个结合区由抗体分子的 V_L 和 V_H 两个结构域构成，如图 11-34 所示。字母“V”表示这一段蛋白质序列是高度变化的。抗体分子的其他结构域叫做“C”区，即稳定区，它在不同的抗体分子之间没有太大变化。

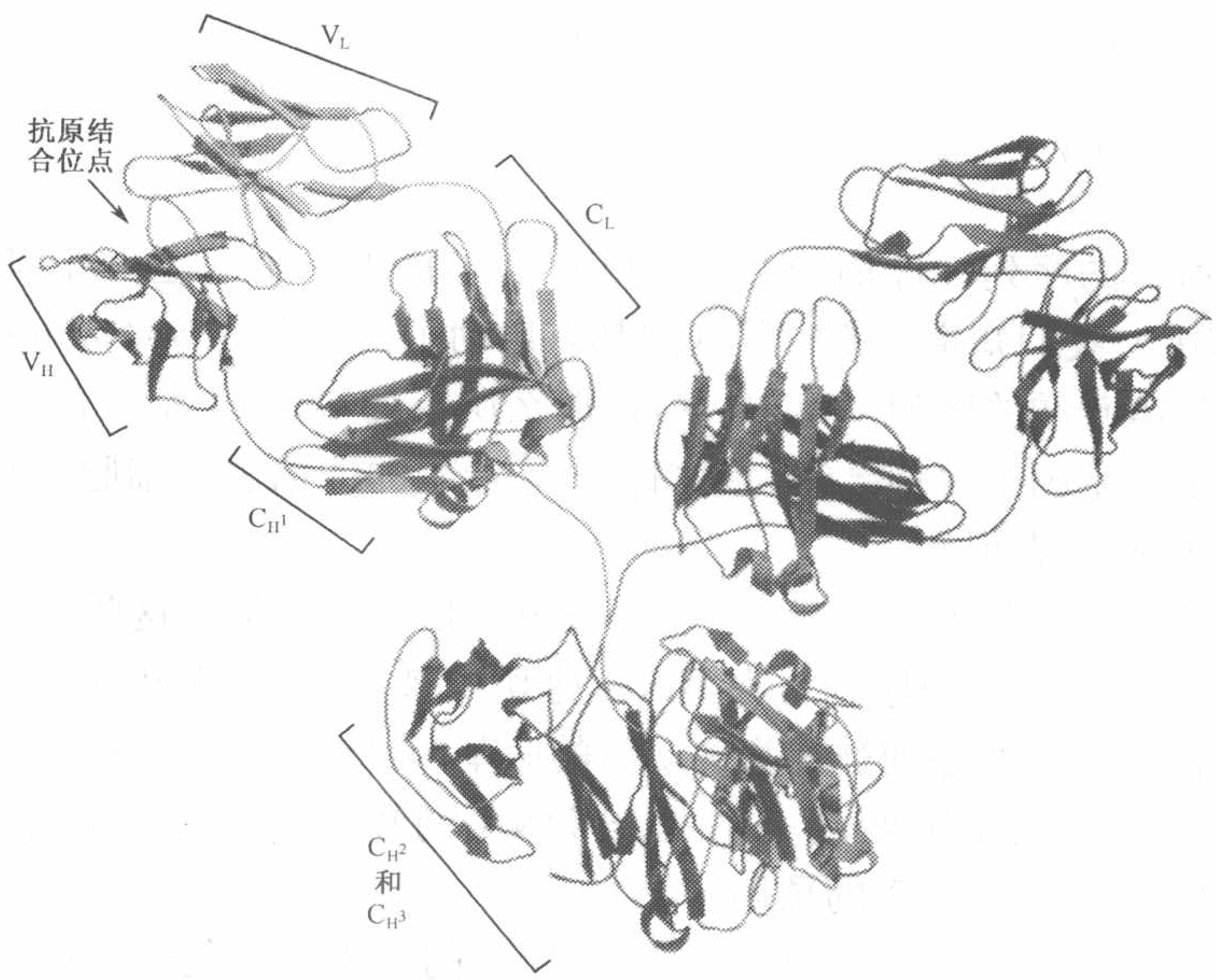


图 11-34 一个抗体分子的结构。两条轻链用粉色显示，两条重链用蓝色表示。可变区和稳定区的标记仅位于抗体分子的左边。注意抗原结合区形成于 V_L 和 V_H 结构域之间表面。（Harris L. J. et al. 1998. J. Mol. Biol. 275：861-872.）图形通过 MolScript、BobScript 和 Raster 3D 软件制作。

如图 11-35a 所示，编码抗体轻链（来自小鼠）的基因组区域叫做 kappa 基因座，这个区域有 300 个基因片段，可以编码不同版本的轻链 V_L 蛋白区。另外有 4 个基因片段编码一个短的蛋白质区域，叫 J 区，其后是编码 C_L 结构域的一个单独编码区。通过我们下面描述的机制，V（D）J 重组可将 V 区和 J 区的 DNA 任意配对融合。所以，单从这一基因组区域重组产生的抗体轻链就有 1200 种变化，然后通过 RNA 剪切将这些基因片段与 C_L 编码区结合在一起（第 13 章）。

编码抗体重链的基因片段装配情况与此相似，不过它还有一个额外的基因单元，叫 D（多样性）片段（图 11-35c），`所以重链的情况更加复杂。例如，小鼠的一个特定重链基因组有 100 多种 V 区、12 种 D 区和 4 种 J 区，组装这个基因的 V（D）J 重组就会产生 4800 多种不同蛋白质序列。由于具有功能的抗体可由任意的重链和轻链配对而成，所以在轻链和重链基因位点进行的重组对抗体蛋白结构的影响具有放大效应。

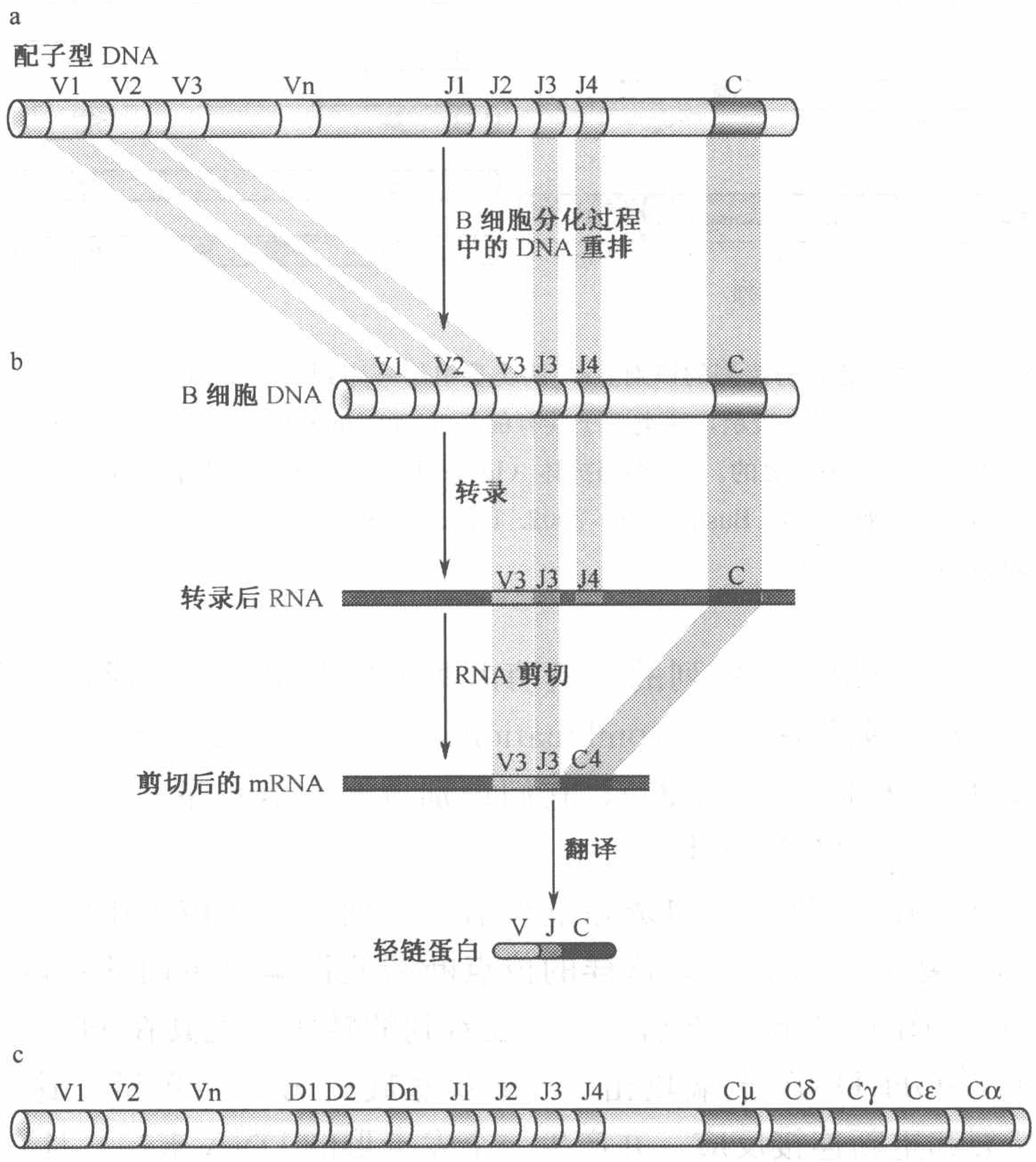


图 11-35 V(D)J 重组过程概述。上图所示为生成抗体蛋白轻链的步骤。a. 细胞中（配子型 DNA）未经 V（D）J 重组的部分轻链 DNA 的遗传结构。b. B 细胞发育过程中发生在两个特异性基因片段（V3 和 J3）间的重组。这是发生在不同 B 细胞前体中很多种重组事件中的一种，重组的基因座转录后通过 RNA 剪切（第 13 章）并列在一起形成连续的基因片段，该 mRNA 翻译产生轻链蛋白。c. 更加复杂的重链遗传区域的图例，它带有额外的“D”基因片段和多种类型的稳定区片段（Cμ、Cγ 等）。（来源：Bushman F. 2002. Lateral DNA transfer，p. 345，Fig. 11. 3. © 2002. Cold Spring Harbor Laboratory Press.）

V（D）J 重组中的早期活动是通过一种类似于转座子切除的机制实现的

被 V（D）J 重组所装配的基因片段两端的序列叫做重组信号序列（recombination signal sequence）。这些信号序列都有两个高度保守的序列基序（motif），一个长 7bp（7mer），另一个长 9bp（9mer）（图 11-36），这些基序能被重组酶结合（如下所述）。重组信号序列分为两类，一类是 7mer 和 9mer 的基序中间为 12bp 的序列；另一类是两个基序中间为 23bp 的序列（图 11-36a）。重组常发生在一对信号序列之间，其中一个有 12bp 的间隔区，另一个有 23bp 的间隔区。这些成对的信号序列以反向重复的结构排列在要被连起来的 DNA 片段两端（图 11-36b）。

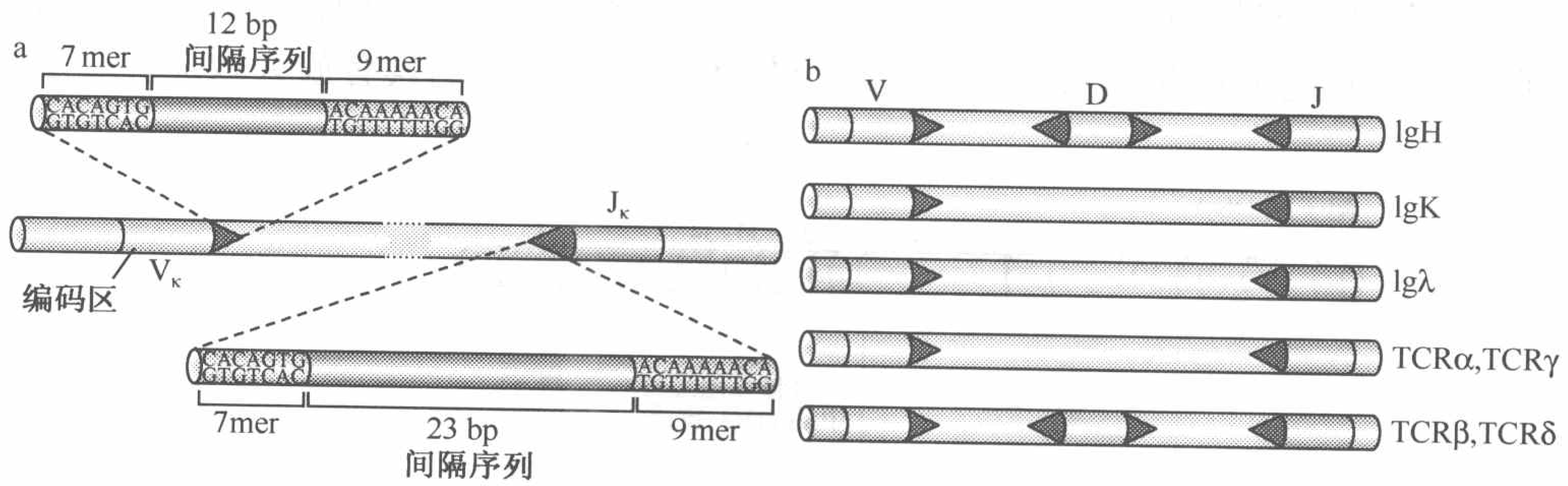


图 11-36　V（D）J 重组中重组信号序列的识别。 a. 两类重组信号序列（RSS）的近观图，12bp 的间隔序列由蓝色显示，23bp 的间隔序列由绿色表示，两类因子共有的 7mer 和 9mer 的序列因子由浅绿色显示，间隔区核苷酸序列并不重要，而长度是至关重要的。b. 编码抗体（Ig 基因）和 T 细胞受体蛋白（TCR 基因）的遗传区域中 RSS 重排的几个例子。（来源：a. Bushman F. 2002. Lateral DNA transfer，p. 345，Fig. 11. 3. © 2002 Cold Spring Harbor Laboratory Press.）

负责识别和断裂重组信号序列的重组酶由两个蛋白质亚基组成，分别叫 RAG1 和 RAG2 [RAG 是重组激活基因（recombination activating gene）的缩写]。这些蛋白质的作用方式与转座酶相似（图 11-37），它们识别重组酶信号序列，然后将两位点配对构成一个蛋白质-DNA 联合复合体。

然后，复合体内的 RAG1 蛋白诱导重组信号序列和将要被重排的基因片段连接处的 DNA 单链的断裂（图 11-37a）。这样的位点断裂使得编码蛋白的片段具有了一个自由的 3′-OH 末端（图 11-37b）。然后，与以前看到的转座子尤其在 Hermes 通路中，见图 11-21，这个 3′-OH DNA 末端攻击 DNA 双螺旋的另一条单链。这个攻击导致了 DNA 的双链断裂和重新连接反应，并产生一个发卡形的 DNA 末端。具有 DNA 发卡末端的序列是蛋白质编码片段，而重组信号序列末端现在是正常的 DNA 双链断裂（图 11-37c）。用同样的机制，每个重组的 DNA 片段都形成一个 DNA 发卡结构。

一旦突触复合体上的两个 DNA 序列被切开，又被 RAG 重组酶变成发卡形状后，细胞内的 DNA 修复蛋白接管任务，并完成剩下的重组反应（图 11-37d）。两个蛋白质编码片段上的 DNA 发卡末端必须被打开，而且这些末端还必须被连接到一起，细胞内的非同源末端连接蛋白（第 9 章）参与了这一反应。有趣的是，DNA 的连接通常伴随

着若干核苷酸的增减，这里的增加与我们讲 Tc*1*/*mariner* 转座子切除时在原靶位点上留下“足迹”的现象类似。这些增加的额外成分为所生成的蛋白质分子序列增加了多样性。成对断裂的重组信号序列也在重组过程中连接了起来，结果产生了一个环形的 DNA 分子，而通常会被细胞丢弃。

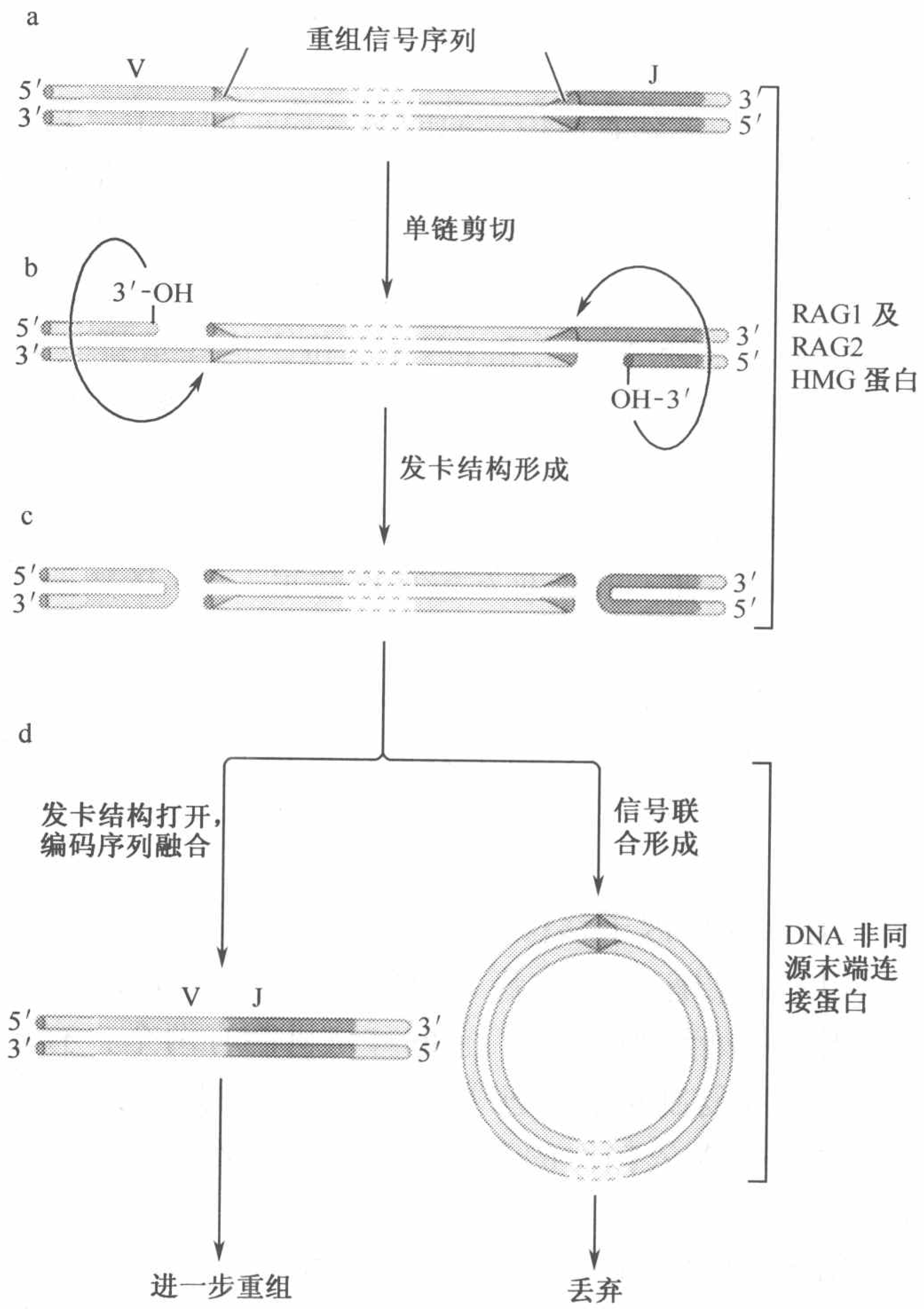


图 11-37　V (D) J 重组途径：发生断裂的机制类似于转座子的切除机制。重组酶催化信号序列末端的单链断裂后产生了游离的 3′-OH，然后每个3′-OH开始攻击相对链形成一个发卡形的中间体（图 11-23b）。随后发卡结构水解并连接形成 V 和 J 区的一个编码联合体，携带重组信号序列的两个末端也连成一个信号联合体，前者可以继续进行重组，后者则被丢弃。（来源：Bushman F. 2002. Lateral DNA transfer, p. 348, Fig. 11. 6. © 2002 Cold Spring Harbor Laboratory Press.）

初始化 V（D）J 重组的 DNA 断裂与转座子切除这两种机制间的相似性是非常明显的。事实上，重组信号序列与转座子末端的末端反向重复序列也很相似，而且 RAG1 蛋白据有一些同 DDE 转座酶蛋白家族相似的序列。事实上，基因组分析近期发现一个叫作 Transib 的转座子家族。它有可能是 RAG1 和重组信号序列的共同来源。这些现象和很多其他现象一起为我们提供的强有力的证据表明，V(D)J 重组这一高等动物免疫

系统的关键特征是由 DNA 转座进化而来的。这个结论说明了可转座因子在细胞基因组进化过程中的关键作用。

小结

尽管 DNA 通常被看作是承载着遗传物质的非常稳定的分子，它也经受着很多种的重排作用。两类遗传重组——保守重组和转座是造成大多数 DNA 重排的原因。

保守重组发生在 DNA 的特定序列因子上。重组酶蛋白识别这些序列因子，对其进行 DNA 单链切断和重新连接，使包含重组位点的 DNA 片段发生重排。有 3 种常见的重排类型：DNA 插入、DNA 丢失和 DNA 倒位。这些重排具有很多功能，如病毒感染时将自己的基因组插入到宿主细胞中，分解 DNA 多聚体，以及改变基因表达。

DNA 上重组位点的结构和 DNA 构筑蛋白的参与决定了某种特定重组反应的发生。构筑蛋白可使得 DNA 片段弯曲，并且对发生在该特定 DNA 区域的重组反应有很大的影响。

有两种保守位点特异性重组酶家族，它们都是通过一个蛋白质-DNA 共价中间体断裂 DNA 的。对于丝氨酸重组酶而言，这个共价连接是通过一个丝氨酸残基活性位点；而对于酪氨酸重组酶而言，则是一个酪氨酸残基活性位点。酪氨酸重组酶的结构揭示出了重组机制的细节。

转座是一类将可移动遗传因子移位到新位点的重组反应，该遗传因子叫转座子。主要有 3 类转座子：DNA 转座子、类病毒反转录转座子和聚腺苷酸反转录转座子。DNA 转座子在转座周期中始终为 DNA，它的移位机制有两个：有切除后转座子中间体参与的剪切-粘贴机制和复制机制。两类反式转座子的移动都要通过一个 RNA 中间体，它们的移位需要 RNA 依赖性的 DNA 聚合酶（又叫反转录酶）和一个重组酶蛋白。

所有物种的基因组中都有转座子，它们可能占据了全部 DNA 序列的一大部分，同时它们还是突变和基因组重排的一个主要原因。转座过程往往受到调控，以确保转座子不会对宿主细胞的基因组造成太大的干扰。通常可以观察到的两种调控机制是对转座子拷贝数的控制和对新插入位点的选择。

最后，一个类似转座的机制被用于其他类型的 DNA 重排反应。最重要的例子是 V(D)J 重组反应，它负责在脊椎动物免疫系统的发育中对基因片段进行装配。

（汪浩　译　冯小黎　胡松年　苏夜阳　校）

参考文献

书

Bushman F. 2002. *Lateral DNA transfer: Mechanisms and consequences.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Craig N.L., Craigie R., Gellert M., and Lambowitz A.M., eds. 2002. *Mobile DNA II.* American Society for Microbiology, Washington, D.C.

位点特异性重组

Chen Y. and Rice P.A. 2003. New insight into site-specific recombination from FLP recombinase-DNA structures. *Annu. Rev. Biophys. Biomol. Struct.* **32:** 135–159.

Grindley N.D.F., Whiteson K.L., and Rice P.A. 2006. Mechanisms of site-specific recombination. *Annu. Rev. Biochem.* **75:** 567–605.

Hallet B. and Sherratt D.J. 1997. Transposition and site-specific recombination: Adapting DNA cut-and-paste mechanisms to a variety of genetic rearrangements. *FEMS Microbiol. Rev.* **21:** 157–178.

Smith M.C. and Thorpe H.M. 2002. Diversity in the serine recombinases. *Mol. Microbiol.* **44:** 299–307.

Stark W.M., Boocock M.R., and Sherratt D.J. 1992. Catalysis by site-specific recombinases. *Trends Genet.* **8:** 432–439.

转座

Haren L., Ton-Hoang B., and Chandler M. 1999. Integrating DNA: Transposases and viral integrases. *Annu. Rev. Microbiol.* **53:** 245–281.

Plasterck R. 1995. The Tc*1/mariner* transposon family. *Curr. Top. Microbiol. Immunol.* **204:** 125–143.

Prak E.T.L. and Kazazian H.H., Jr. 2000. Mobile elements in the human genome. *Nat. Rev. Genet.* **1:** 134–144.

Rice P.A. and Baker T.A. 2001. Comparative architecture of transposase and integrative complexes. *Nat. Struct. Biol.* **8:** 302–307.

Smit A.F.A. 1999. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* **9:** 657–663.

Williams T.L. and Baker T.A. 2000. Transposase team puts a headlock on DNA. *Science* **289:** 73–74.

转座及V(D)J重组

Gueguen E., Rousseau P., Duval-Valentin G., and Chandler M. 2005. The transpososome: Control of transposition at the level of catalysis. *Trends Microbiol.* **13:** 543–549.

Oettinger M.A. 2004. Hairpins at split ends in DNA. *Nature* **432:** 960–961.

V(D)J重组

Fugmann S.D., Lee A.I., Schockett P.E., Villey I.J., and Schatz D.G. 2000. The RAG proteins and V(D)J recombination: Complexes, ends, and transposition. *Annu. Rev. Immunol.* **18:** 495–527.

Gellert M. 2002. V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu. Rev. Biochem.* **71:** 101–132.



第3篇

基因组的表达

本篇概要

- 第 12 章 转录机制
- 第 13 章 RNA 剪接
- 第 14 章 翻译
- 第 15 章 遗传密码

第 3 篇的内容涉及了解基因所面临的重要挑战之一——基因是如何表达的。或者说，以核苷酸线形序列的形式储存在多聚核苷酸链中的信息是如何转换成为多肽链中的氨基酸线形序列的？遗传信息由基因向蛋白质的流动以及信息流的单向性概念，即人们所熟知的中心法则，是由 Francis Crick 于 1958 年阐明的。

中心法则指明，“信息”一旦传入了蛋白质，便不能再传出。信息从核酸传递到核酸，或者从核酸传递到蛋白质，都是可能的；但是从蛋白质传递到蛋白质，或者从蛋白质传递到核酸，则是不可能的。这里的信息是指核酸中的碱基序列或蛋白质中的氨基酸残基序列的精确确定。

第 12～15 章是按照将基因中的信息复制到 RNA，即信使 RNA（messenger RNA），再将信使 RNA 解码成为多肽链的信息流向程编排的。核苷酸序列的信息由 DNA 传递到 RNA 的过程，即转录（transcription），是第 12 章的主题。

一种多亚基的分子机器，即 RNA 聚合酶，在 DNA 双螺旋中形成一个移动的“泡”。DNA 在这个转录泡的前缘是松散的，而在转录泡的尾端则又重新缠绕成为螺旋结构。RNA 聚合酶以转录泡内短暂分离的两条 DNA 链中的一条为模板，三个与 DNA 合成极其相似的化学反应中通过碱基配对逐渐合成互补的 RNA 链。尽管在所有细胞中合成 RNA 的基本酶非常相似，但真核生物转录机器的其他部分要比原核生物复杂得多。这个复杂性的提高为调控更多的真核基因提供了必不可少的额外控制点。同样，DNA 中决定转录起始的序列（启动子）和决定转录停止的序列（终止子）在真核生物中也更复杂。

在原核生物中，信使 RNA 一旦合成，便立即作为蛋白质合成的模板而使信息流向下一个阶段。而真核生物则不然：转录的 RNA 产物必须经过一系列的成熟事件才能成为信使 RNA。第 12 章描述了其中两个事件：在 5′端添加所谓的“帽子”结构和在 3′端添加多聚腺苷酸（poly-A）尾巴。

最引人注目的加工事件称为信使 RNA 剪接（mRNA splicing），第 13 章对此进行了描述。真核生物细胞的基因常常被一个，有时是多个非蛋白质编码的片段（即内含子）所打断。当这样的基因转录成 RNA 的时候，这些内含子必须去除，而使蛋白质编码片段（即外显子）能够彼此连接在一起，成为一条连续的蛋白质编码序列。第 13 章描述了负责高准确度去除内含子的精密分子机器。

第 3 篇以第 13 章和第 14 章的翻译（translation）过程作为终结。翻译是以核苷酸

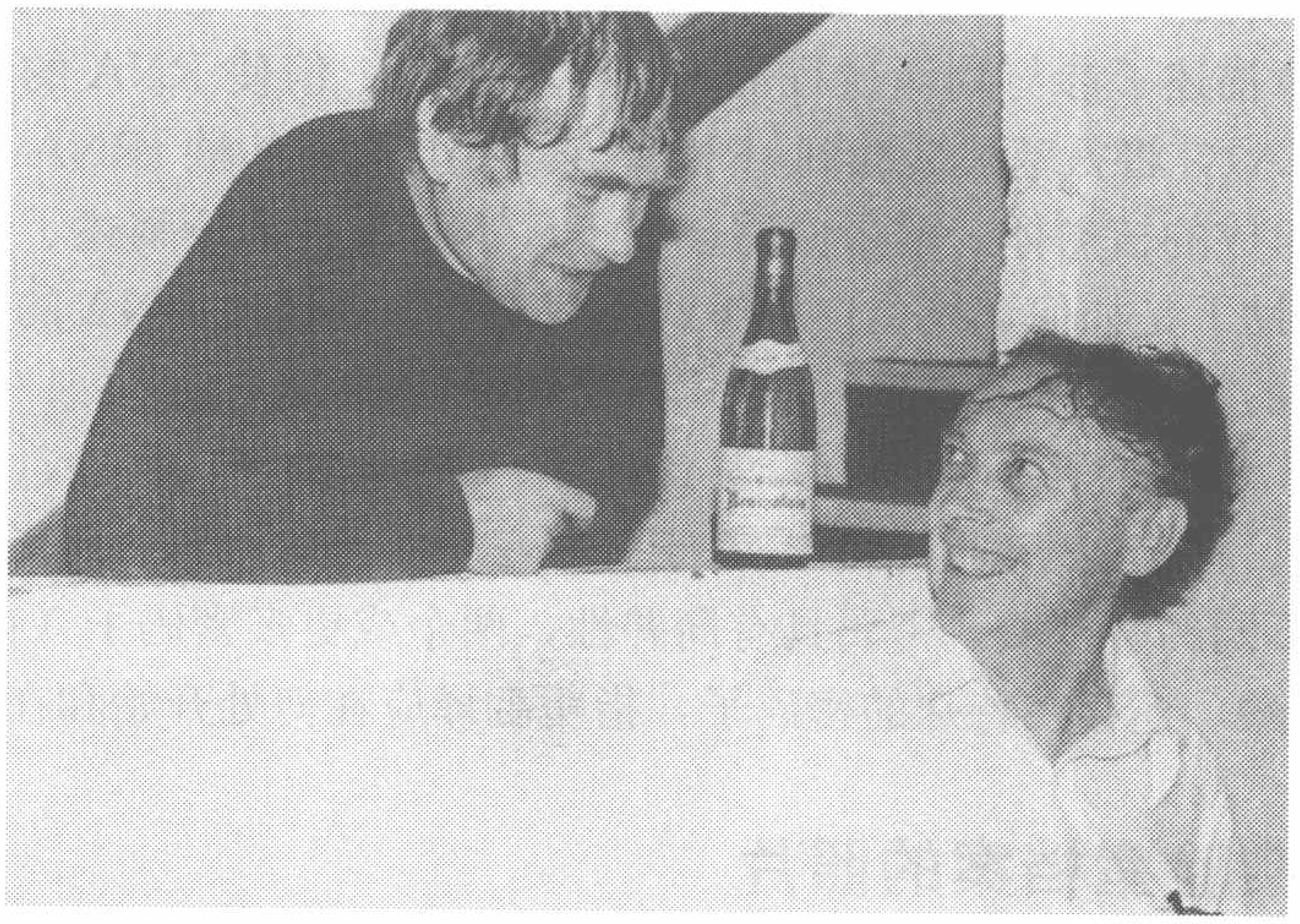
序列形式储存在信使 RNA 中的遗传信息用于指导氨基酸有序地掺入到蛋白质多肽链中的过程。第 14 章描述了参与翻译过程的 4 个主要组分：信使 RNA 的编码序列，转运器分子（即 tRNA），将氨基酸装载到 tRNA 转运器上的氨基酰 tRNA 合成酶，以及由 RNA 和蛋白质组成的核糖体，即蛋白质合成工厂本身。本章还阐述了这 4 个组分在许多辅助因子的帮助下是如何识别蛋白质编调控把某一特定信使 RNA 的核苷酸编码转换成其蛋白质产物的正确氨基酸顺序这样一个奇异的过程。

最后，第 15 章描述了阐明遗传密码（genetic code）的经典实验，同时列出了翻译这些密码的法则。核苷酸序列的信息是基于一种三个字母的密码，而蛋白质序列的信息则是基于 20 个不同的氨基酸。密码具有简并性，两个或更多密码子（在大多数情况下）编码同一种氨基酸。还存在特殊的密码子，指明翻译应在何处开始和在何处结束。

来自冷泉港实验室档案的照片



Richard Roberts，1977 年染色质研讨会。Roberts 的主要研究集中于限制性内切核酸酶的功能和多样性（第 21 章），但他也是“割裂基因”的共同发现者之一，为此他与 Phillip Sharp 于 1993 年一起获得了诺贝尔生理及医学奖。这里展示的和他在一起的科学家是（由左到右）：Yasha Gluzman，肿瘤病毒学家；Ahmad Bukhari，致力于研究噬菌体 Mu 转座（第 11 章）；James Darnell，他的工作集中在基因调节中的信号转导方面（第 17 章）。



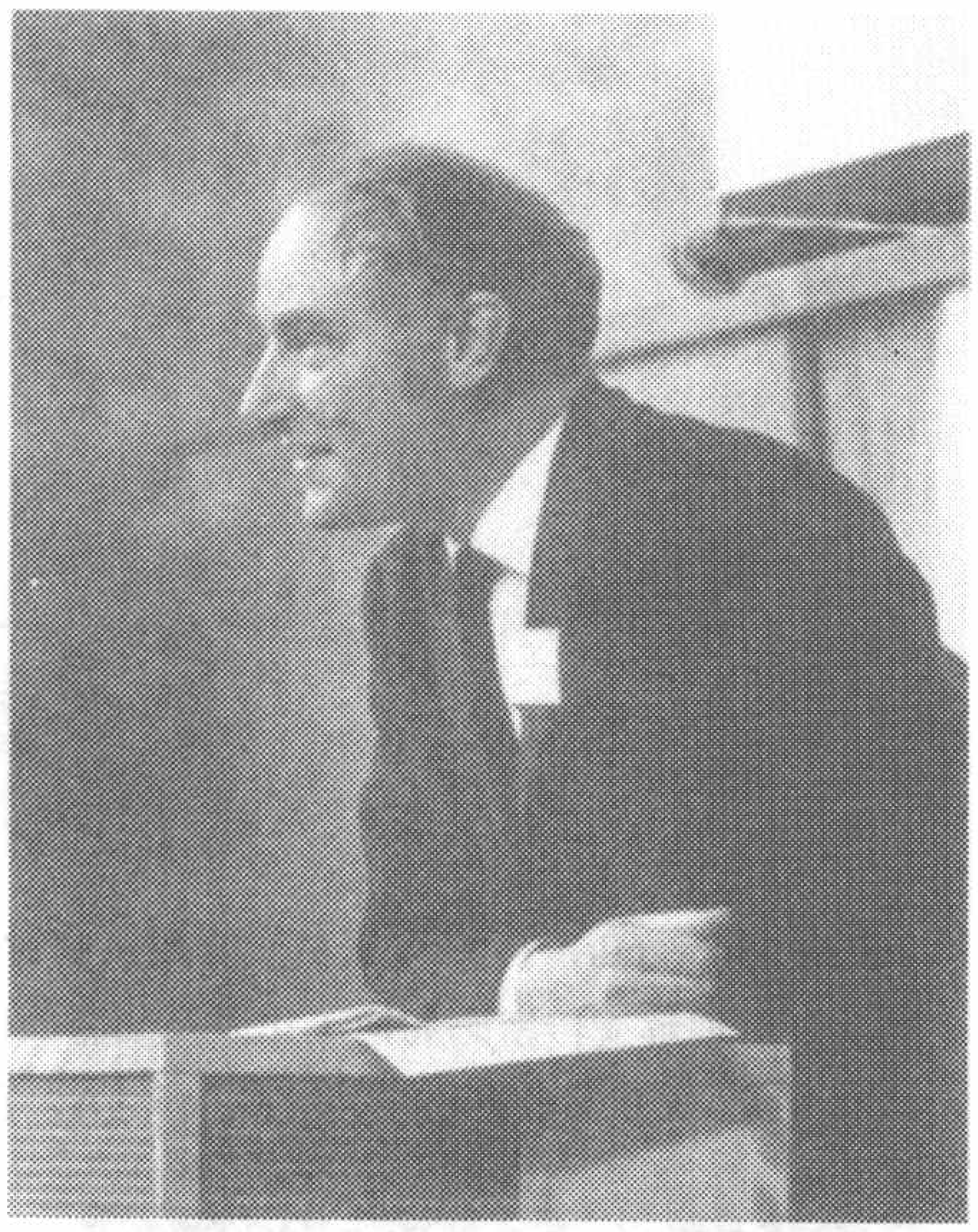
Sydney Brenner 和 James Watson，1975 年突触讨论会。 Brenner，图中和 Watson 在一起，为 mRNA 和遗传密码本质的发现（第 2 章和第 15 章）做出了贡献，而他于 2002 年与别人共同获得诺贝尔奖则是因为他为发育生物学研究建立了秀丽新小杆线虫（*C. elegans*）模式系统（第 22 章）。



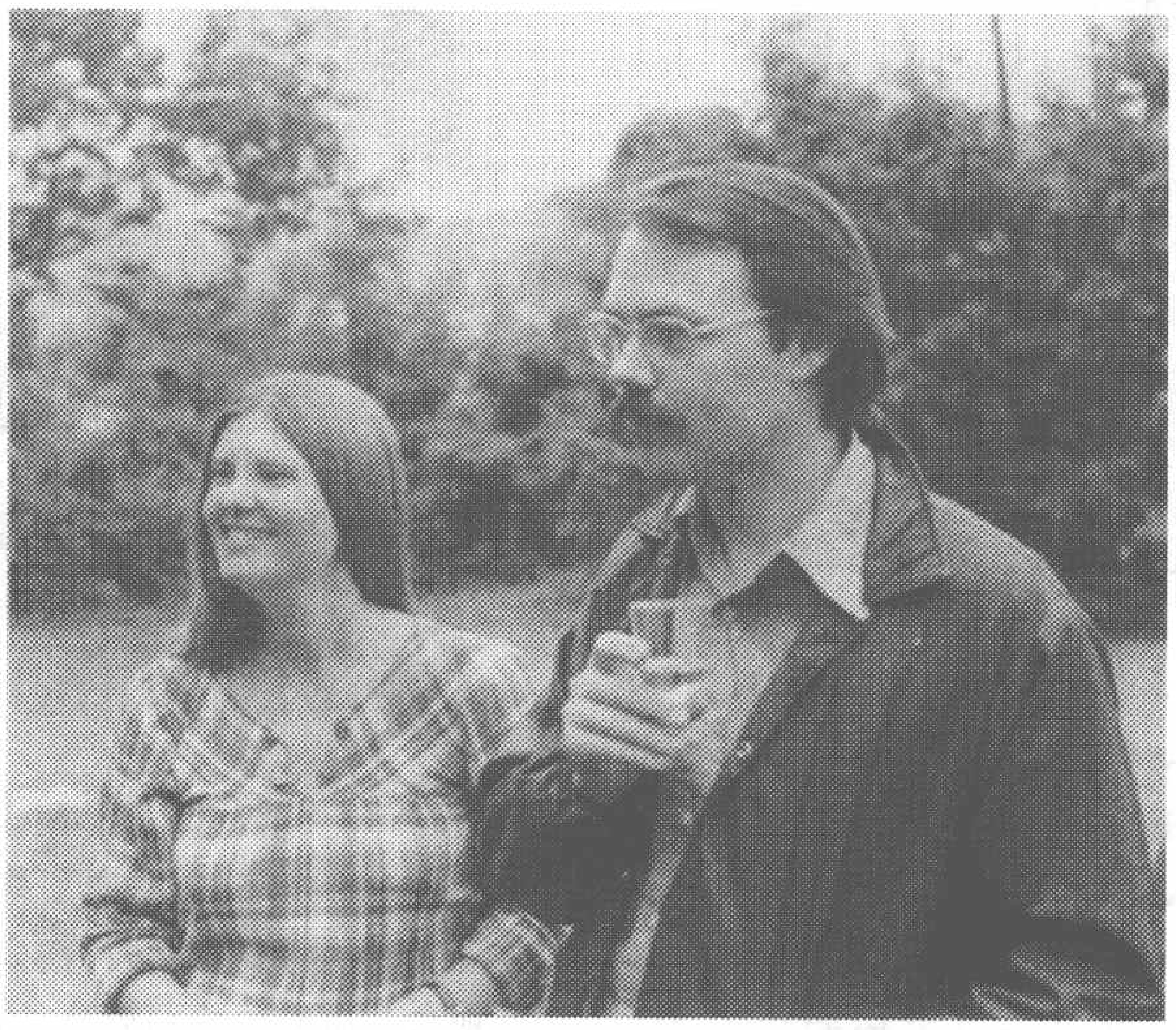
Robert Roeder，1998 年转寻机制讨论会。 Roeder 发现了三个真核 RNA 聚合酶，Pol Ⅰ、Ⅱ、Ⅲ，纯化了这些酶以及从它们各自的启动子起始转录所需的其他因子（第 12 章）。左边那个带着质疑眼神的是 Camilo Parada，当时是 Roeder 实验室的博士后。



Roger D. Kornberg，1977 年染色质讨论会。早期研究核小体结构（第 7 章），Kornberg 于 2006 年因 RNA 聚合酶结构（第 12 章）的研究获得诺贝尔化学奖。他父亲是 Arthur Kornberg，第 98 页有其相片。



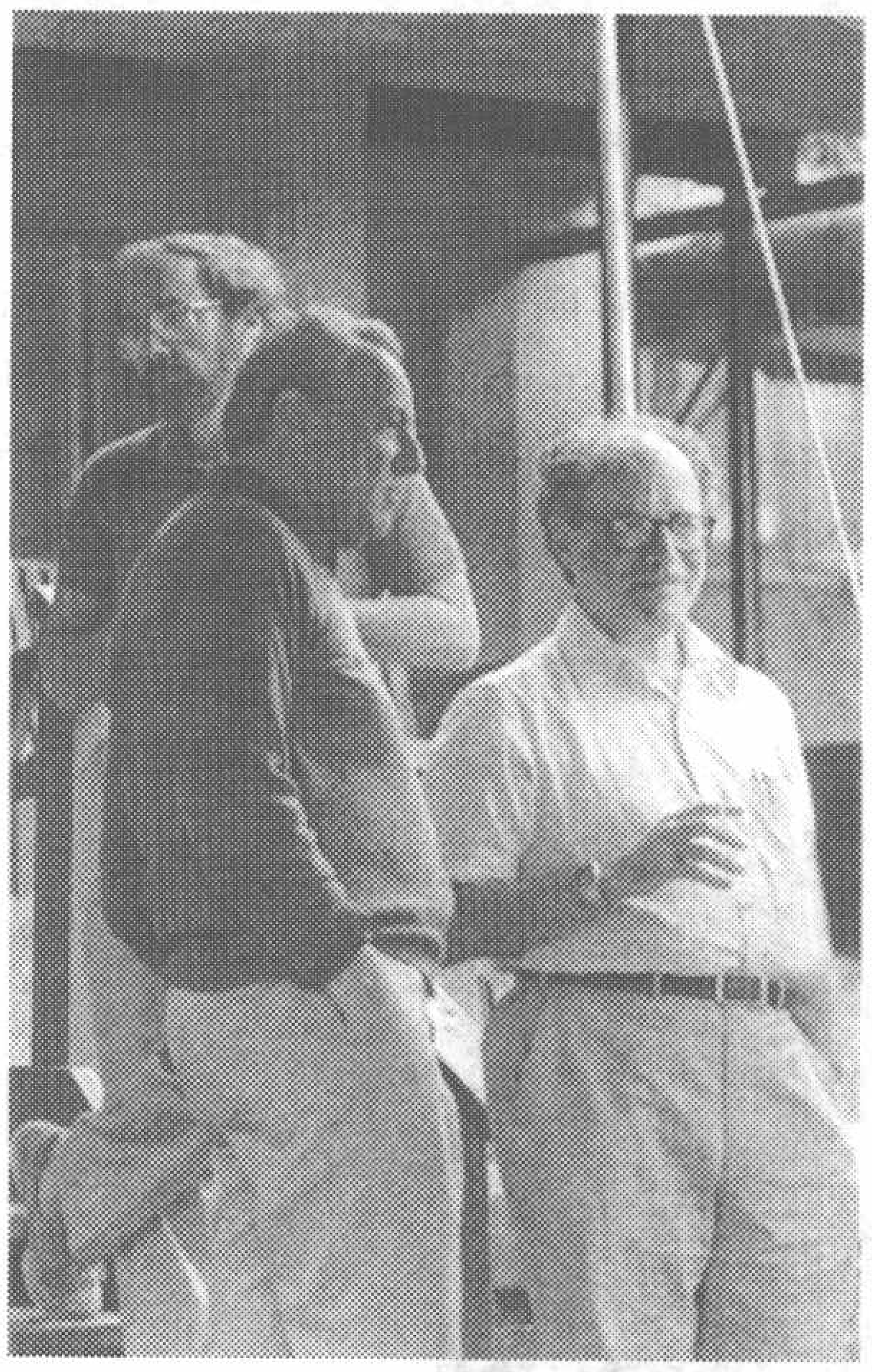
Francis Crick，1963 年大分子的合成与结构讨论会。除了在解决 DNA 结构中的贡献之外，Crick 还是分子生物学领域早期发展的智能驱动力。他的“转配器假说”（发表在 RNA Tie Club newsletter 上）预言了存在将 RNA 的遗传密码翻译成蛋白质的氨基酸序列所必需的分子。后来发现 tRNA 就是行使这一功能的分子（第 14 章）。



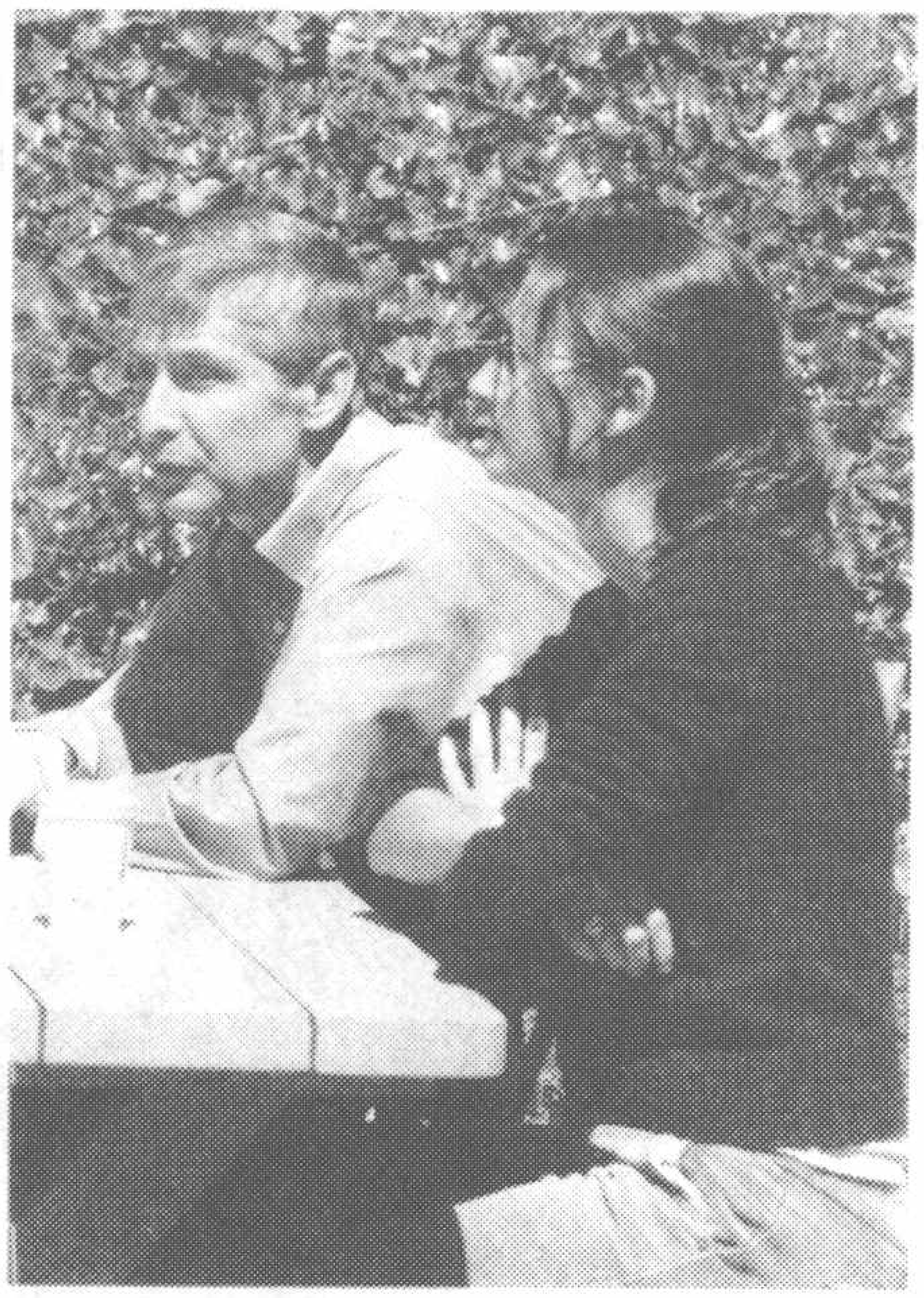
Phillip Sharp，1974 年肿瘤病毒讨论会。Sharp 和 Richard Roberts 共同获得了 1993 年的诺贝尔生理及医学奖，因为他们发现很多真核细胞的基因是“割裂”的，也就是说，它们的编码区域是被一段一段的非编码 DNA 打断的。这些非编码区域是通过“剪接”而从 RNA 拷贝中去除的（第 13 章）。图中为 Sharp 和夫人 Anne。



Paul Zamecnik，1969 年蛋白质合成机制讨论会。 Zamecnik 发展了体外蛋白质合成系统，已证明该系统对理解遗传密码如何工作以及细胞如何制造蛋白质是非常重要的（第 2 章和第 15 章）。他还和 Mahlon Hoagland 一起发现了 tRNA，它是蛋白质合成过程中的一个关键成分（第 14 章）。



David Baltimore、Francois Jacob 和 Walter Gilbert，1985 年发育分子生物学讨论会。 Baltimore 和 Howard Temin 共同发现了以 RNA 为模板合成 DNA 的反转录酶（第 11 章）。Jacob 与 Jacques Monod 提出了基因表达如何被调节的基本模型（第 16 章），还提出了 DNA 复制如何被调节的一个模型（第 8 章）。Gilbert 为基因调节的“Jacob 和 Monod 模型”的某些方面提供了生物化学证明，他还发明了一种用于 DNA 测序的化学方法（第 20 章）。他们分别于 1975 年（生理及医学）、1965 年（生理及医学）和 1980 年（化学）与其他人一起获得了诺贝尔奖。



David Allis 和 Emily Bernstein，2004 年表观遗传学讨论会。 Allis 第一个发现了修饰组氨酸的酶，Tetrahymena 的组氨酸乙酰转移酶（第 7 章）。自从这个酶被发现后，研究组氨酸修饰范围及其对基因表达的影响的全新领域成长起来了。图中 Allis 与他实验室的博士后 Bernstein 在一起，后者曾是 Greg Hannon 实验室的研究生，在那里她发现了参与 RNAi 的 Dicer 酶（第 18 章）。

第12章 转录机制

本章概要

•

RNA 聚合酶和转录周期

•

细菌的转录周期

•

真核生物的转录

•

由 RNA 聚合酶Ⅰ和Ⅲ催化的转录

在此之前，我们一直在考察基因组的维持，也就是说，遗传物质是如何组织、保护和复制的。现在，我们将问题转到遗传物质如何表达，即 DNA 的一串碱基是如何指导 RNA 和蛋白质的合成，后者行使细胞功能和定义细胞特性。在后面的几章里，我们将描述负责基因表达的基本过程：转录、RNA 加工以及翻译。

转录在化学和酶学上与 DNA 的复制（第 8 章）非常相似。二者都通过酶的作用合成一条与 DNA 模板互补的核酸链。当然，二者之间存在一些重要的差别，最主要的就是转录得到的新链是由核糖核苷酸组成的，而不是脱氧核糖核苷酸（第 6 章）。转录机制区别于复制的其他特征包括以下几点：

- RNA 聚合酶（RNA polymerase，催化 RNA 合成的酶）不需要引物，它能够从头起始转录（但是，如我们将要讨论的，细胞内的转录必须在特定的序列上才能起始）。
- RNA 产物不与模板 DNA 链的碱基保持互补状态。相反，RNA 聚合酶在每个核糖核苷酸添加处的几个核苷酸之后，便将正在延长的链从模板上置换下来（图 12-1）。这一置换对 RNA 行使功能非常关键（例如，最常见的是翻译产生蛋白产物）。而且，由于这一释放紧随合成位点之后，多个 RNA 聚合酶分子可以一个紧接着另一个同时转录同一个基因。这样，一个细胞就可以在短时间内在一个基因（或其他 DNA 序列）上合成大量的转录产物。重要的是，当 RNA 产物紧跟在每个行进的 RNA 聚合的后面从 DNA 模板上解离下来时，两条 DNA 链重新配对（图 12-1）。
- 转录尽管是非常精确的（每添加 10 000 个核苷酸会发生一次错误），但还是不如复制更为精确（每 10 000 000 个核苷酸发生一次错误）。这一区别反映出转录过程缺乏广泛的校正机制，尽管确实存在两种 RNA 合成的校正形式。

较转录而言，细胞更关注复制的精确性是有道理的。DNA 是储存遗传物质的分子，DNA 复制是遗传物质得以传递的过程。复制过程中发生的任何错误都会很容易引起灾

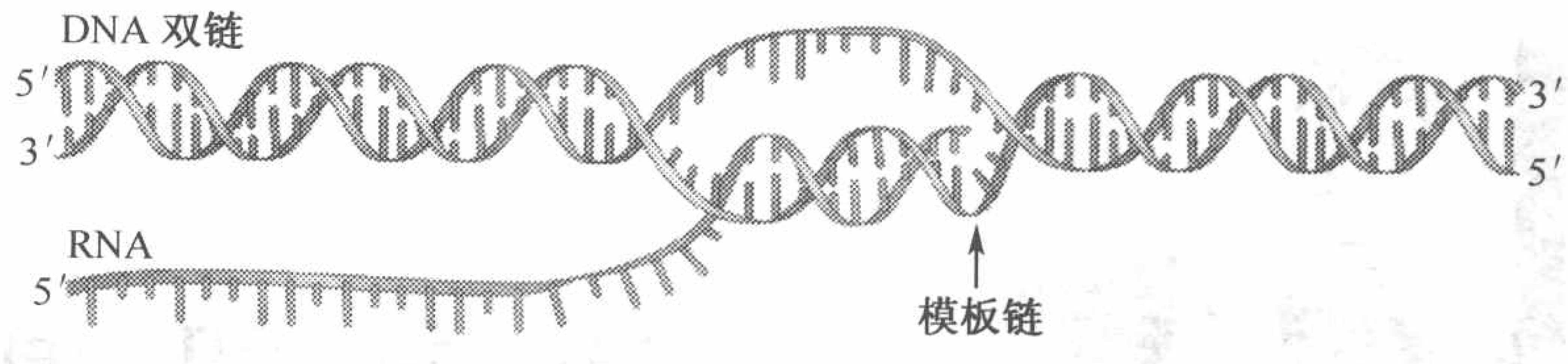


图 12-1 DNA 转录成为 RNA。本图所示为 DNA 双螺旋是如何解旋和 RNA 链是如何在模板上合成的（RNA 聚合酶未在图中显示）。本图还显示了 RNA 转录产物是如何在合成点，几个核苷酸后从 DNA 模板上解离下来，以及 DNA 链如何重新配对。图中，转录从左到右进行。

难性的后果：它将永远存在于这个个体的基因组中，并传递给它的后代。相反，转录仅仅产生一些暂时性的拷贝，而且通常从同一个转录区域会得到几个拷贝。因此，转录过程中发生一个错误，除了在许多暂时性的副本中产生一个有缺陷的副本外，几乎不会造成任何其他的损害。

除了上述机制的区别，DNA 复制和转录之间还存在一个重要的差异，反映出这两个过程的不同目的。转录仅是有选择性地复制基因组的特定部分，并从任何特定的部分产生一个到几百个，或者甚至上千个拷贝。相反，复制必须将整个基因组全部拷贝，并且在每次细胞分裂时进行一次（并且只进行一次）（第 8 章）。转录区域的选择并不是随机的：每个转录区域通常包括一个或多个基因，而且在每个转录区域的首端和尾端都存在特定的 DNA 序列指导转录的起始和终止。

不仅基因组的不同部分转录的程度不同，而且转录哪一部分、转录范围有多大都是可以调控的。这样，在不同的细胞中，或者同一细胞的不同时段，可能转录不同组别的基因。例如，某个人的两个遗传特征相同的细胞在很多情况下会转录不同组别的基因，从而导致那两个细胞在特征和功能上的差异（如一个可以是肌肉细胞，而另一个是神经细胞）。或者，一个特定的细菌细胞会依其生长的培养基的不同，而转录不同组别的基因。这些调控的问题将在第 4 篇予以解答。

RNA 聚合酶和转录周期

RNA 聚合酶具有不同形式，但有许多共同特性

从细菌到人类，在所有细胞里，RNA 聚合酶在本质上都进行同一种反应。因此来自这些不同物种的酶具有许多共同的特性就不足为奇了，尤其是那些直接参与催化 RNA 合成的部分。从细菌到哺乳动物细胞的 RNA 聚合酶都是由多个亚基构成的（虽然，一些噬菌体和细胞器的确编码一些单亚基的酶来完成相同的任务，框 12-2）。表 12-1 列出了各种情况下亚基的数目和大小，以及不同的酶之间哪些亚基序列是保守的。

表 12-1 RNA 聚合酶的亚基

原核生物		真核生物		
细菌	古细菌	RNAP Ⅰ	RNAP Ⅱ	RNAP Ⅲ
核心（core）	核心（core）	(Pol Ⅰ)	(Pol Ⅱ)	(Pol Ⅲ)
β'	A′/A″	RPA 1	RPB 1	RPC 1
β	B	RPA 2	RPB 2	RPC 2
α^{I}	D	RPC 5	RPB 3	RPC 5
α^{II}	L	RPC 9	RPB 11	RPC 9
ω	K	RPB 6	RPB 6	RPB 6
	[+6 其他的]	[+9 其他的]	[+7 其他的]	[+11 其他的]

注：每列中的亚基均按分子质量降序排列。

来源：Data adapted from Ebright R. H. 2000 J. Mol. Biol. 304：687-698. Fig. 1. p. 688. © Elsevier.

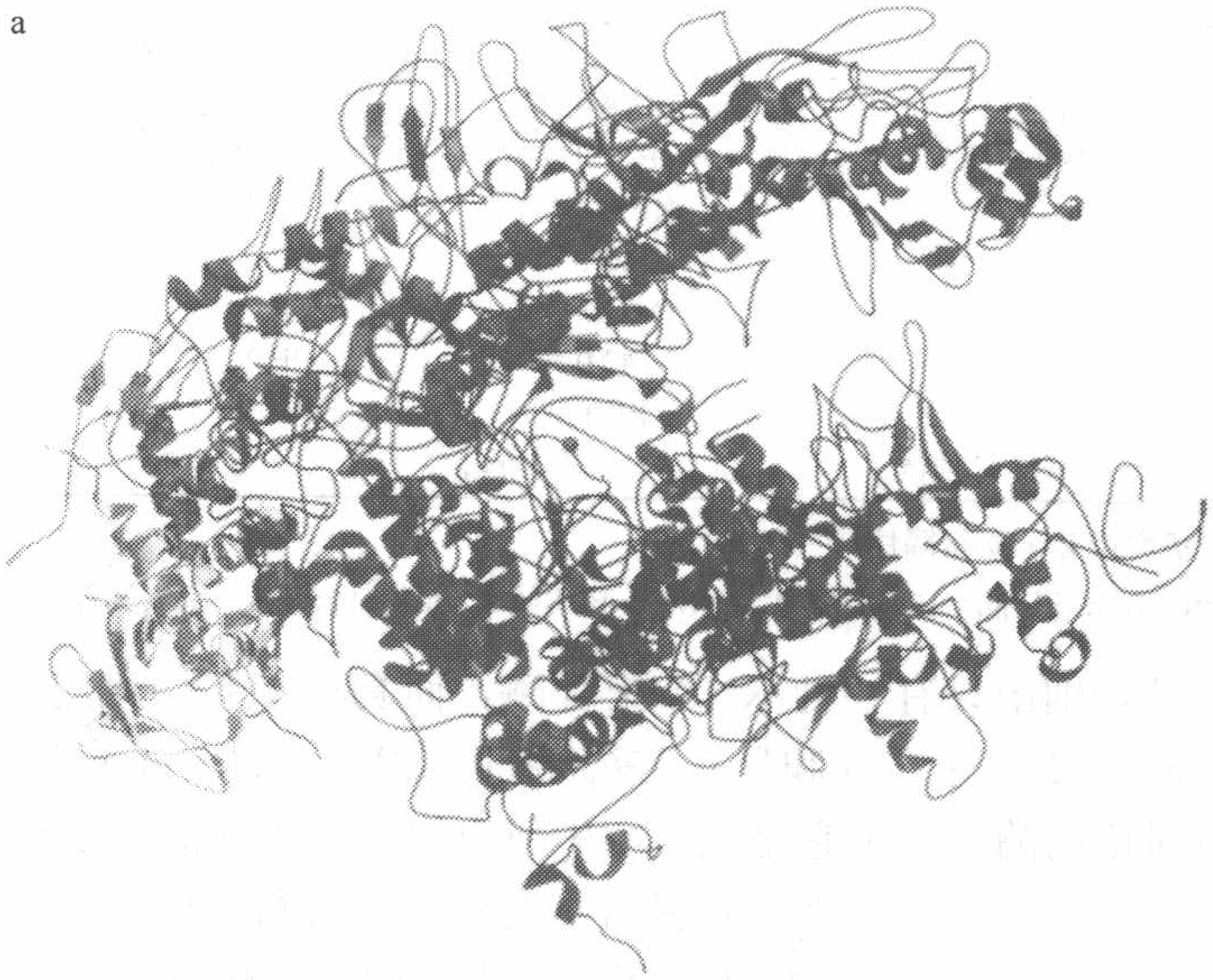
由此表中可见，细菌只有一种 RNA 聚合酶，而真核细胞中则有 3 种：RNA 聚合酶Ⅰ、Ⅱ和Ⅲ（RNA Pol Ⅰ，Ⅱ和Ⅲ）。聚合酶Ⅱ（Pol Ⅱ）是本章后半部分阐述真核转录时将要重点讨论的酶，因为它是这些酶中研究得最为深入的，也是负责转录大多数基因的酶——实际上，是转录几乎所有蛋白质编码基因的酶。聚合酶Ⅰ（Pol Ⅰ）和聚合酶Ⅲ（Pol Ⅲ）分别参与转录专门的 RNA 编码基因。具体来说，Pol Ⅰ转录的是大核糖体 RNA 前体基因，而 Pol Ⅲ转录的是 tRNA 基因、一些小的核 RNA 基因和 5S rRNA 基因。在本章的末尾，我们将再讨论这些酶。

细菌 RNA 聚合酶的核心酶（core enzyme）可以独立合成 RNA，它是由 α 亚基（subunit）的两个拷贝和 β、β'、ω 亚基的各一个拷贝组成的。这个酶与真核生物的聚合酶密切相关（表 12-1)。具体来说，它的两个大亚基，β 和 β'，与 RNA 聚合酶Ⅱ的两个大亚基（RPB1 和 RPB2）同源。其 α 亚基与 RPB3 和 RPB11 同源、ω 亚基与 RPB6 同源。细菌 RNA 聚合酶核心酶的结构与酵母 RNA 聚合酶Ⅱ的核心酶相似。二者在图 12-2中并列显示。后面我们将会描述一些能够阐明这些酶的工作原理的结构细节，这里我们只突显一些共同的特征。

细菌和酵母的核心酶有着共同的总体形态和组织形式；实际上，它们比经亚基序列比较所预期的结果更为相似。二者内部靠近活性位点的部分尤为如此，而外周部分的相似性则差一些。这些相似点和相异点的分布是合理的：酶的内部参与在 DNA 模板上合成 RNA（所有生物体都一样)；然而，酶的许多外周部分参与与其他蛋白质的相互作用，而这些在原核和真核细胞中是不同的。

总的来说，每一个酶的形状都像一只蟹爪。这让我们回想起第 8 章中描述的 DNA 聚合酶的“手”形结构（图 8-5)。蟹爪的两个钳子主要是由各个酶的两个最大亚基构成（细菌的酶是 β 和 β' 亚基，真核生物的酶是 RPB1 和 RPB2 亚基)。酶的活性位点是由这两个亚基的一些区域共同组成的，位于钳子基部被称为“活性中心裂隙”的区域（图 12-2)。活性位点的工作原理依据双金属离子催化添加核苷酸的机制，该机制在所有的聚合酶中均如此（第 8 章)。然而，RNA 聚合酶的活性位点含有一个紧密结合的 Mg^{2+} 离子，第二个离子是在合成周期中由每个新的核苷酸带入，和焦磷酸一起释放。

a



b

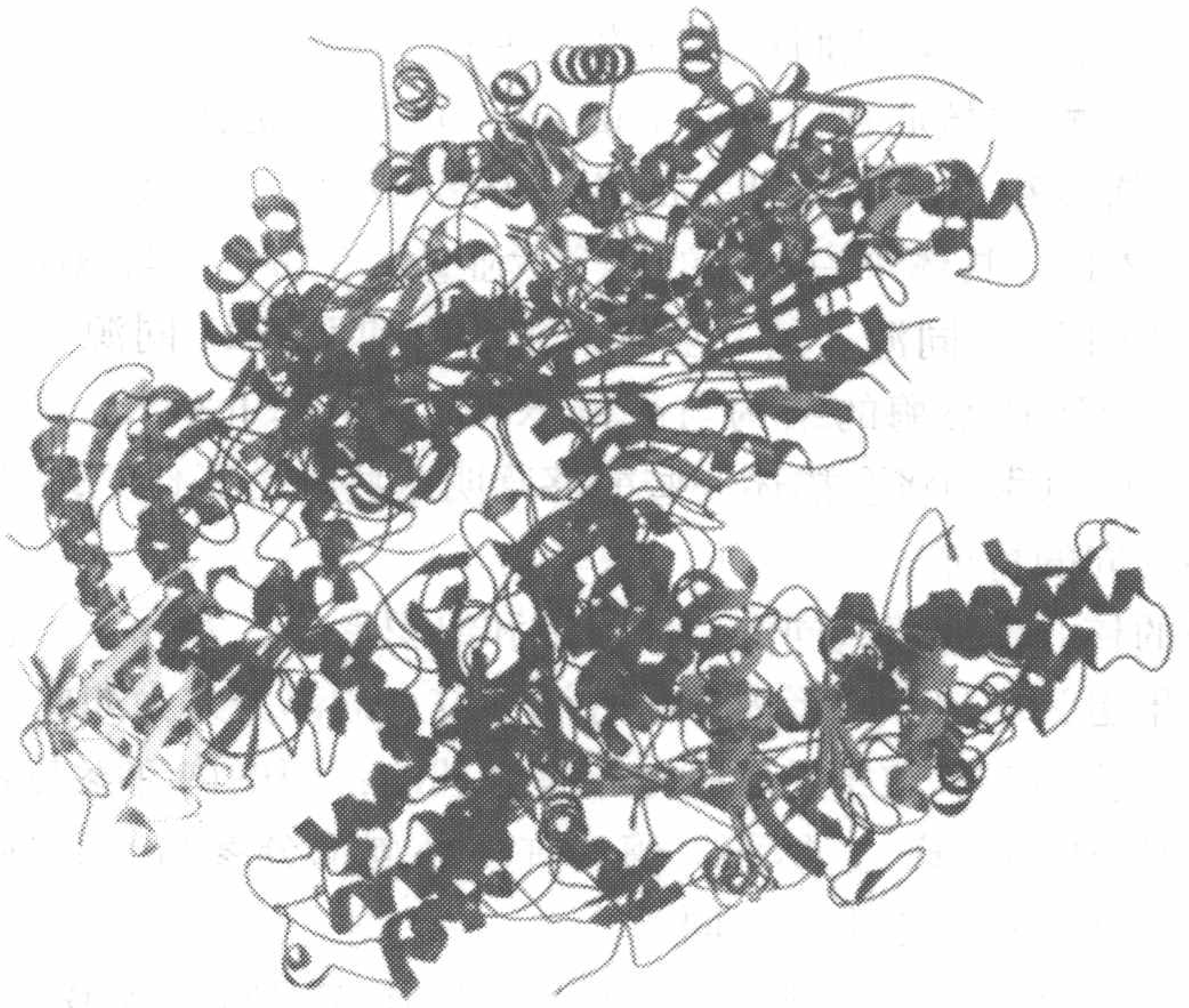


图 12-2 原核生物和真核生物 RNA 聚合酶晶体结构（crystal structure）的比较。 a. 栖热水生菌（*T. aquaticus*）RNA 聚合酶的核心酶结构。各个亚基以不同的颜色表示：β 亚基为紫色，β′亚基为蓝色，两个 α 亚基分别为黄色和绿色，ω 亚基为红色以红球表示的 Mg^{2+} 在 a、b 中均注明了活性位点。（Seth Darst，洛克菲勒大学，私人通信）。b. 酿酒酵母（*S. cerevisiae*）RNA 聚合酶Ⅱ的结构。各个亚基以不同的颜色表示，以表明与细菌中相应亚基的关系（表 12-1）。因此，RPB1 和 RPB2 分别为紫色和蓝色，RPB3 和 RPB11 分别为黄色和绿色，RPB6 为红色（Cramer P.，et al. 2001. Science. 292：1863）。图像使用 MolScript、BobScript 和 Raster 3D 等软件制作。

有多种通道允许 DNA、RNA 和核糖核酸进出这些酶的活性中心裂隙。我们会在后面谈到转录机制的时候再进行讨论。

由 RNA 聚合酶进行的转录是由一系列步骤完成的

为转录一个基因，RNA 聚合酶要进行一系列定义明确的步骤，这些步骤分为 3 个阶段：起始（initiation）、延伸（elongation）和终止（termination）。在这里及图 12-3 中，我们总结了每个阶段的基本特征。

起始 启动子（promoter）是最初结合 RNA 聚合酶的 DNA 序列与所需的转录起始因子一起结合）。启动子-聚合酶复合体一旦形成就发生结构改变，以利起始过程继续进行。和复制的起始一样，DNA 在转录将要开始的部位解旋，碱基对分开，产生一个单链 DNA 的“泡”。与 DNA 复制相同，转录总是从 5′端向 3′端进行。也就是说，新的核糖核苷酸添加在延伸链的 3′端。然而，与 DNA 复制不同的是，只有一条 DNA 链作为模板合成 RNA 链。因为 RNA 聚合酶以固定的方向与启动子结合，所以从一个特定的启动子，总是同一条链被转录。

启动子的选择决定了哪一段 DNA 被转录，这也是对转录进行调控的主要步骤。也就是说，是否启动某个基因的转录，是细胞在特定时期调控特定蛋白质合成的主要手段。

延伸 一旦 RNA 聚合酶已经合成了一小段 RNA（10 个碱基左右），转录便进入了延伸阶段。在延伸阶段，RNA 聚合酶除催化 RNA 的合成外，还执行着多项重要任务。它解开前方的 DNA，又使其后的 DNA 重新复性；在移动的同时，逐步解开正在延长的 RNA 链与模板的配对；而且还执行校正的功能。相反，在复制过程中，需要几个不同的酶来执行这些相似的功能。

终止 一旦 RNA 聚合酶转录了整个基因（或整组基因），它必须停下来并释放 RNA 产物（并从 DNA 上解离下来）。这一步骤称为终止。在某些细胞中，存在着特定的、已研究清楚的终止序列；而在其他一些细胞中，是什么使聚合酶停止转录并从模板上分离，还不十分清楚。

转录起始包括三个定义明确的步骤

转录周期（transcription cycle）的第一个阶段——起始又可以分为一系列定义明确的步骤（图 12-3）。第一步是聚合酶与启动子初步结合形成所谓的闭合式复合体（closed complex），这时 DNA 还保持着双链状态，聚合酶结合在螺旋的一个面上。起始阶段的第二步，闭合式复合体经过转变而成为开放式复合体（open complex），DNA 双链在起始位点周围大约 14bp 的距离内分开以形成转录泡。在起始阶段的下一步，启动子逃离，聚合酶进入转录起始阶段。

DNA 的展开使模板链得以释放。头两个核糖核苷酸被带入活性位点，排列在模板链上，并彼此结合。通过这种方式，后续的核糖核苷酸结合到正在延长的 RNA 链上。最初 10 个左右核糖核苷酸的合成是一个效率相当低的过程，在这一阶段，聚合酶经常释放出很短的转录物（每一个都短于 10 个左右的核苷酸），然后再重新开始合成。在这

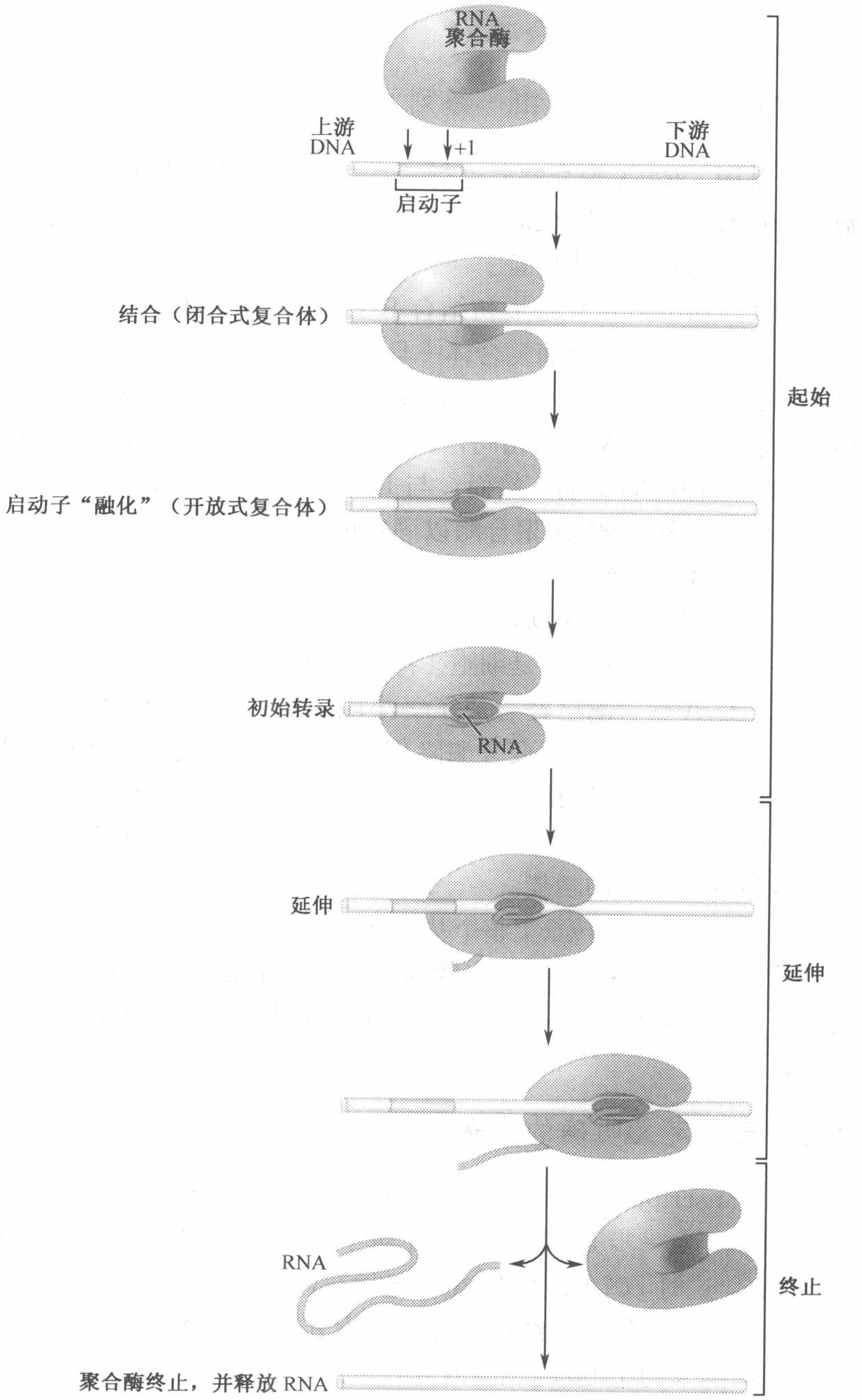


图 12-3 转录周期的几个阶段：起始、延伸和终止。图中显示了转录周期的常规步骤。所示特征为细菌和真核生物转录共同拥有的。起始、延伸和终止所需要的其他因子并未在这里显示，但将在后面的正文中进行描述。编码 RNA 链开端的 DNA 核苷酸称为转录起始位点，指定为“+1”位置。转录行进方向的序列称为起始位点的下游。类似地，起始位点前面的序列称为上游序列。当提到上游序列的某一特定位置时，将给予一个负值，下游序列则给予正值。

个阶段，聚合酶-启动子复合体称为起始转录复合体（initial transenbing complex）。一旦某个聚合酶超过了 10bp，就可以说它逃离了启动子。这时，它已经形成了一个稳定的三重复合体（stable ternary complex），包括酶、DNA 和 RNA。这是向延伸阶段的转变。

在本章的后续部分，我们将更为详细地对转录周期进行描述——首先是细菌，然后是真核系统。

细菌的转录周期

细菌的启动子在强度和序列上是多样的，但是具有某些明确的特征

原则上，细菌的核心 RNA 聚合酶能够在 DNA 分子上的任何一点开始转录，这点可以在体外用强化的核心酶来验证。但在细胞内，聚合酶只在启动子处起始转录。是一个称为 σ 的起始因子的加盟，使得核心酶（$\alpha_2\beta\beta'\omega$）转变为只在启动子处起始转录。该酶的这种形式称为 RNA 聚合酶的全酶（holoenzyme，图 12-4）。

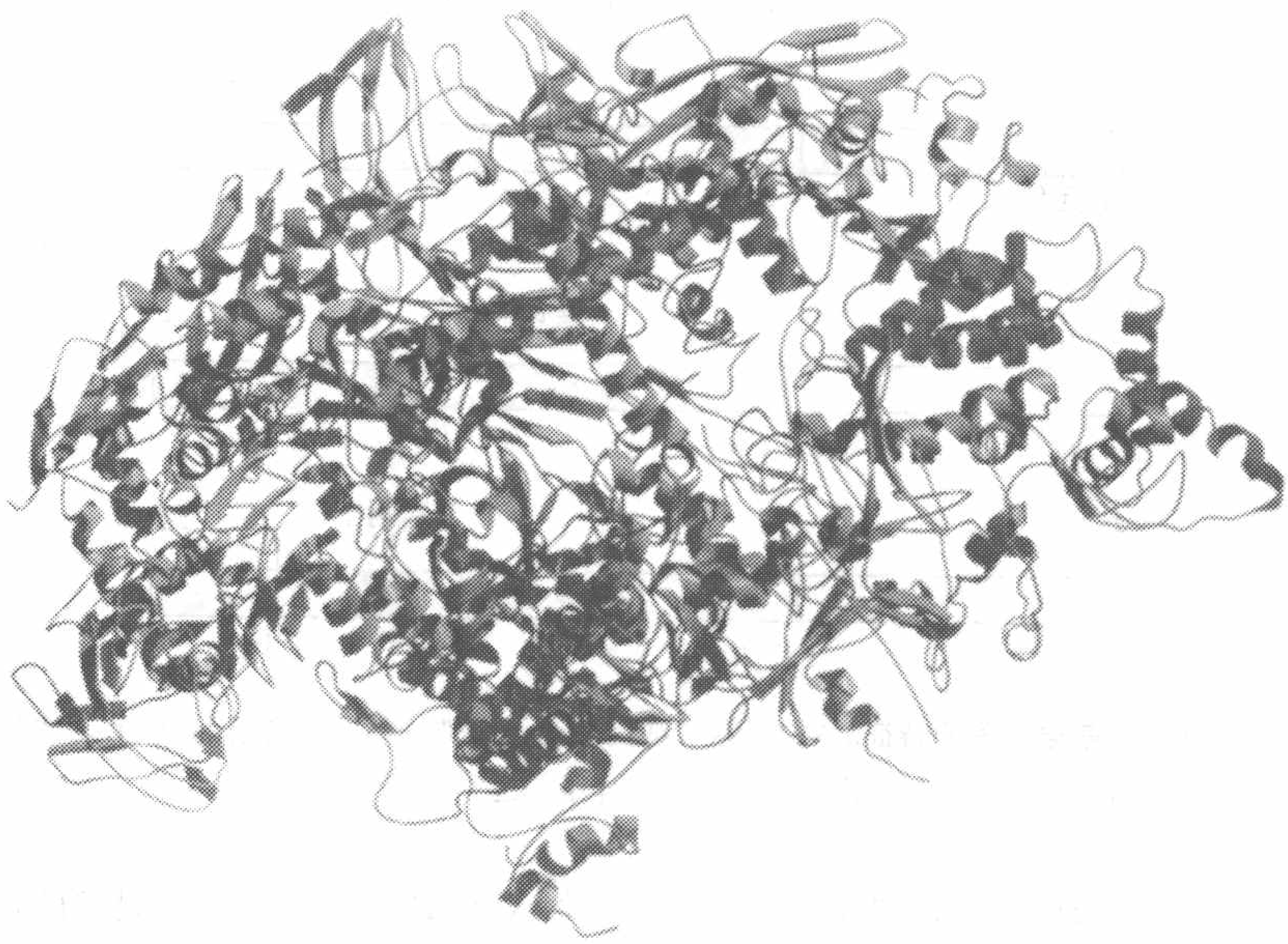


图 12-4　栖热水生菌（*Thermus aquaticus*）RNA 聚合酶的全酶。灰色所示为核心酶（与图 12-2 的 a 图中所示的是同一个酶），紫色所示为 σ^{70} 亚基（区域 2、3、4，见图 12-6）。右边为区域 2，顶部为区域 3，底部为区域 4。σ 区域 2 和 4 分别识别启动子的 −10 和 −35 区域，稍后将在正文中进行描述。（Murakami K. S.，et al. 2002. Science. 296：1280.）图像使用 MolScript、BobScript 和 Raster 3D 等软件制作。

在大肠杆菌中，最常见的 σ 因子称为 σ^{70}（我们将在第 16 章中讨论其他的可供选择的 σ 因子）。含有 σ^{70} 的聚合酶识别的启动子具有以下共同的特征性结构：两段 6 个核苷酸长的保守序列，被长为 17～19 个核苷酸的非特异序列（图 12-5）所分隔。这两段保

守序列的中心分别位于 RNA 合成起始位点上游约 10bp 和 35bp 处。因此，根据图 12-3 的计数方案（编码 RNA 链开端的 DNA 核苷酸被指定为＋1），这两段序列称为－35 和－10 区域（region）或元件（element）。

虽然绝大多数的 σ^{70} 启动子包含可被识别的－35 和－10 区域，但序列并不一致。通过对许多不同启动子的比较，可以得到一个共有序列（consensus sequence，框 12-1）。共有序列反映了首选的－10 和－35 区域，被最适宜的长度（17bp）所分隔。很少有启动子具有与其一模一样的序列，但多数与其仅有几个核苷酸的差别。

具有与共有序列更近似序列的启动子通常要比那些匹配较差的启动子"更强"。所谓启动子的强度，是指一个启动子在一定时间内可以起始多少个转录物。这一度量标准受到以下因素的影响：启动子最初与聚合酶的结合程度、对异构化作用的支持效率，以及此后聚合酶逃离的难易程度。启动子强度与序列的相关性解释了为什么有许多不同种类的启动子：一些基因的表达水平远远超出其他基因，前者很可能具有与共有序列更为接近的序列。

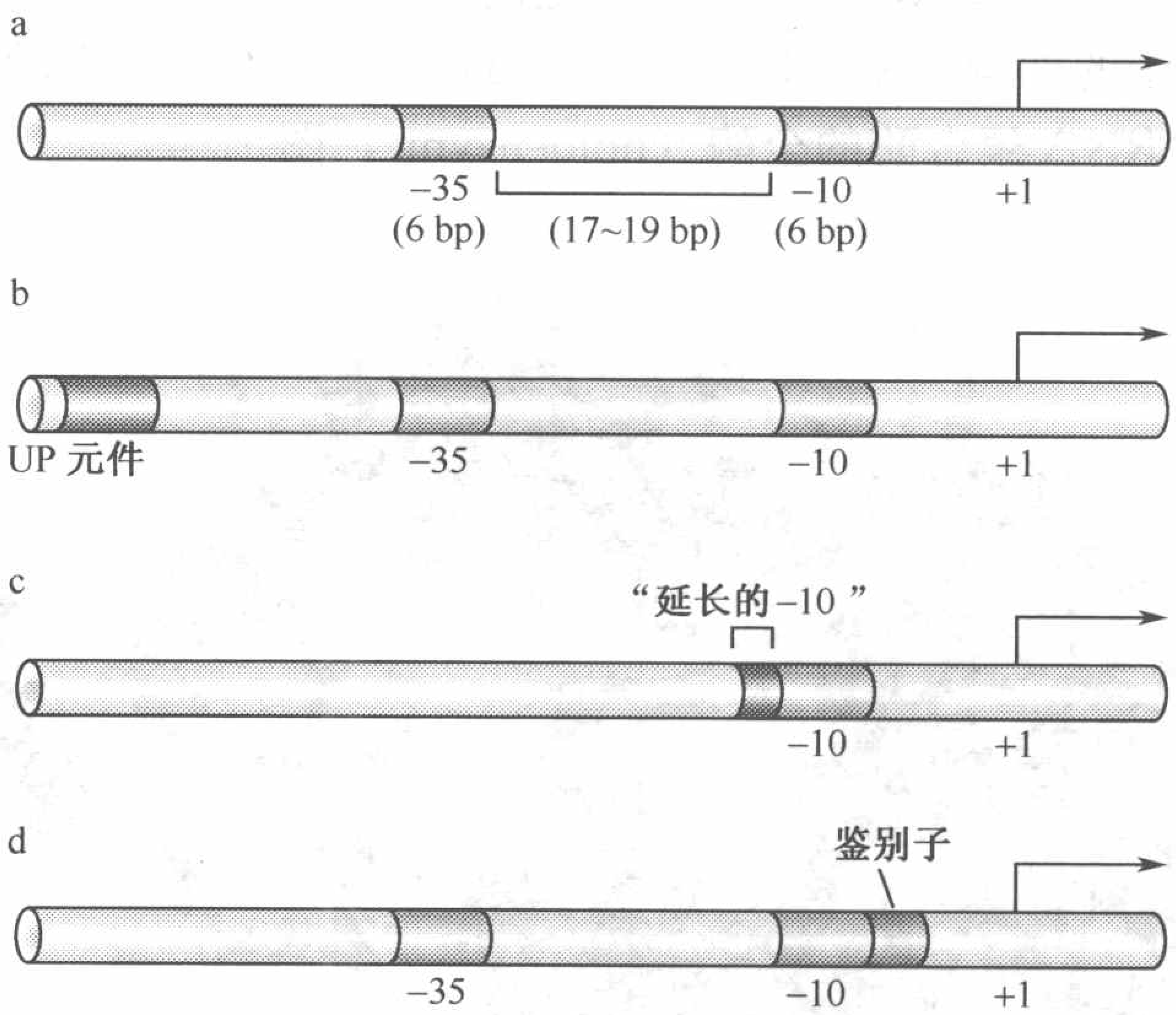


图 12-5　细菌启动子的特征。图中显示了细菌启动子元件的不同组合情况。每个元件对聚合酶的结合和功能所起的作用在正文中有详细描述。

在一些较强的启动子中，如指导核糖体 RNA（rRNA）基因表达的启动子，发现额外与 RNA 聚合酶结合的 DNA 元件，称为 UP 元件（UP-element，图 12-5b）。它通过提供额外的特异性相互作用而增强聚合酶与 DNA 的结合。

另一类型的 σ^{70} 启动子缺乏－35 区域，而是有一个所谓的"延长的－10 区域"元件（图 12-5c）。与标准的－10 区域相比，这一元件在其上游末端有一个额外的短序列元件。聚合酶与这一额外序列元件之间的联系，补偿了－35 区域的缺失。大肠杆菌的 *gal* 基因（该基因产物指导乳糖的代谢，第 16 章）使用的就是这种启动子。

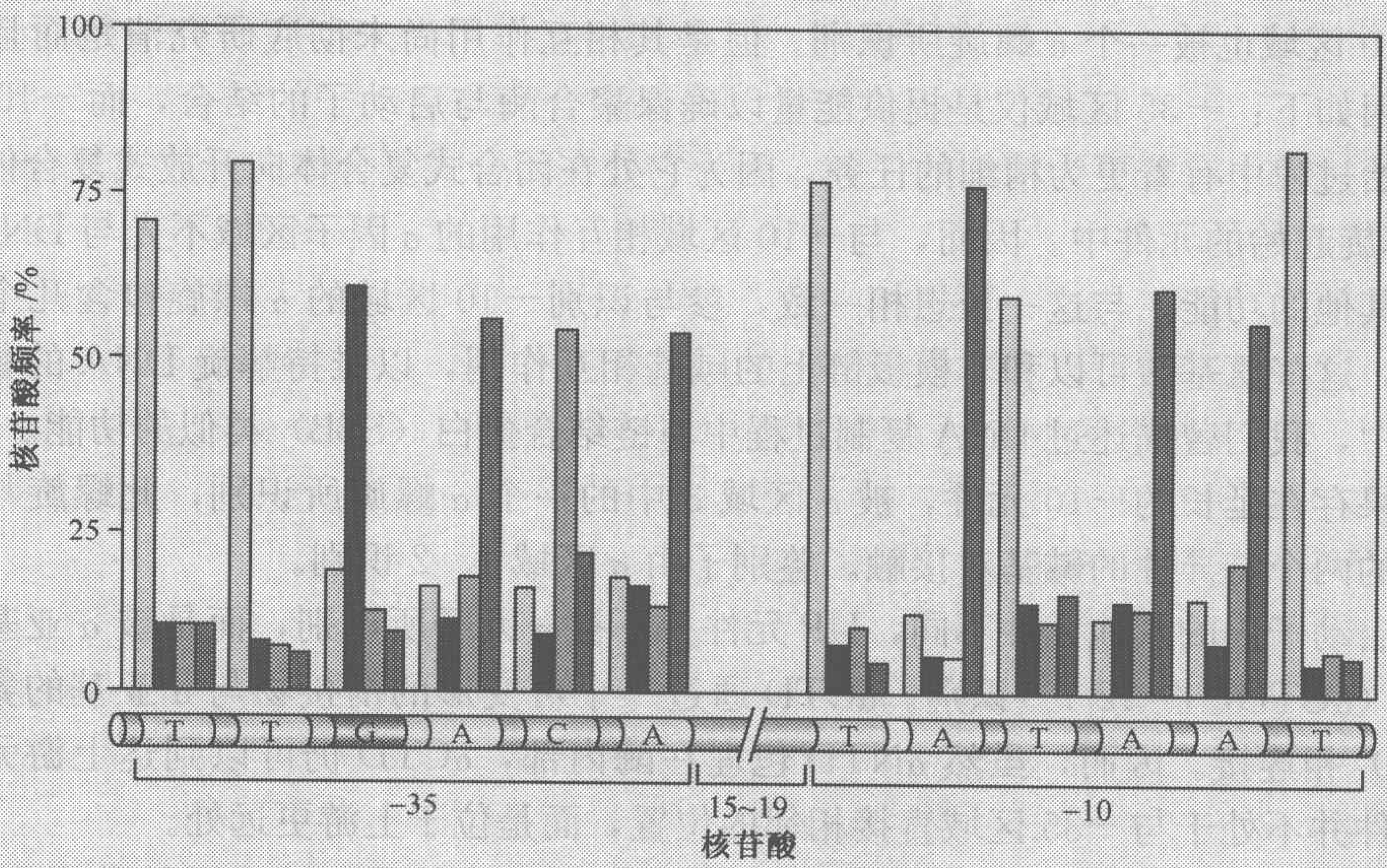
最近，在紧挨着－10 元件的下游发现了一个新的结合 RNA 聚合酶的 DNA 元件。这个新元件称为鉴别子（discriminator），见图 12-5d。辨别子与聚合酶之间相互作用的强度影响酶与启动子复合体的稳定性。

框 12-1　共有序列

被某个特定蛋白质识别的结合位点其 DNA 序列不一定完全一致。类似地，使一个蛋白质具有某种特殊功能的氨基酸序列在不同的蛋白质中可能会有微小的差异。共有序列就是在上述的每种情况中，由不同样本在每一个位置上最常见的核苷酸（或氨基酸）所组成的序列版本。含有 σ^{70} 因子的 RNA 聚合酶识别大肠杆菌启动子的共有序列如图所示（框 12-1 图 1）。这一共有序列是通过对 300 个已知具有 σ^{70} 启动子功能的序列进行比较，并确定在—35 和—10 六元区域中每一个位置上最常见的碱基得到的。该核苷酸就被选为共有序列中该位置上的特定核苷酸；每一个位置上的特定核苷酸出现的相对频率以及其他 3 种核苷酸的频率在图表中进行了描绘。注意，位于—35 和—10 区域之间的 17～19 个核苷酸并不存在具有显著意义的共有序列。

在图例中，每一个单独的启动子序列都已确定，所以比较这些序列并没有什么价值。但是，让我们来考虑一个完全不同的例子：所涉及的 DNA 结合蛋白还没有确定的结合位点。但是，一条染色体的几个区域在其长度范围内的某个位置上含有结合位点。可以使用一种扫描所有这些染色体区域序列的计算机算法来寻找它们共有的潜在结合位点。

当结合位点未知时，找到 DNA 结合蛋白所识别的共有序列的第二种途径是，利用化学方法合成大批随机序列的短 DNA 片段。将目标蛋白与这些 DNA 分子混合在一起，就可以筛选到与目标蛋白结合的 DNA，并测定其序列。对这些序列的比较很容易显示出共有序列，因为每一个片段都非常短。这种方法（常被称为 SELEX）广泛地用于确定未知的 DNA 结合蛋白的结合位点。第 21 章将对 SELEX 进行详细描述。



框 12-1 图 1　启动子的共有序列和共有间距。（来源：Redrawn from Alberts B. et al. 2002. Molecular biology of the cell, 4th edition, p. 308, Fig 6.12. Copyright © 2002. Reproduced by permission of Routledge/Taylor & Francis Books, LLc.）

σ 因子介导聚合酶与启动子的结合

σ^{70}因子可以分为 4 个区域，称为 σ 区域 1～σ 区域 4（图 12-6）。识别启动子－10 和－35 元件的区域分别是区域 2 和 4。

区域 4 的两个螺旋形成一个常见的 DNA 结合基序，称为螺旋-转角-螺旋（helix-turn-helix）。其中一个螺旋插入到－35 区域的大沟（major groove）中并与该区的碱基相互作用；另一个则从大沟的顶部横过，与 DNA 骨架接触。这一结构基序见于许多 DNA 结合蛋白，已经在第 5 章进行了详细的讨论（图 5-20）。具有这一结构基序的例子包括在细菌细胞中发现的几乎所有的转录激活因子与抑制因子（第 16 章）。

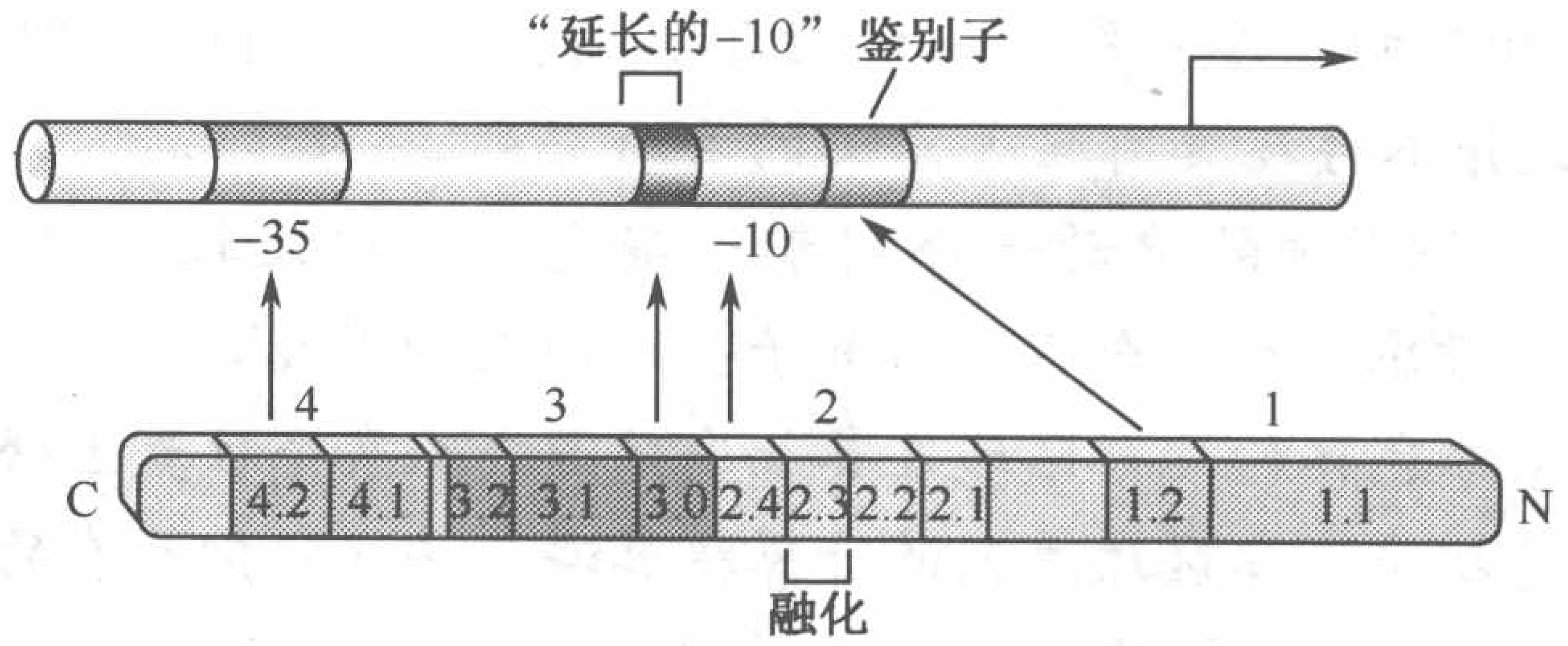


图 12-6 σ 因子的区域。识别启动子特定区域的 σ 因子的那些区域以箭头指示。区域 2.3 负责 DNA 解旋。有关 σ 因子将 RNA 聚合酶核心酶募集到一个标准启动子上的示意图，见图 12-7。（来源：Adapted，with permission，from Young B. A.，et al. 2002. Cell 109：417-420，Fig. 1. © Elsevier.）

－10 区域也被一个 α 螺旋所识别。但是其相互作用尚未彻底研究清楚而且更为复杂，原因如下：－35 区域仅是提供能量以确保聚合酶与启动子的结合，而－10 区域在转录起始过程中有着更为精细的任务，因为它处在闭合式复合体向开放式复合体转变时 DNA 解旋起始的元件中。因而，与－10 区域相互作用的 σ 因子区域不光与 DNA 结合，还参与其他的功能。与这一预想相一致，参与识别－10 区域的 α 螺旋包含几个芳香族氨基酸，这些氨基酸可以和非模板链上的碱基相互作用，以维持解旋 DNA 的稳定。在第 8 章中，我们曾描述过 DNA 复制过程中单链结合蛋白（SSB）类似的功能。

如果存在延长的－10 元件，被 σ 区域 3 中的一个 α 螺旋所识别，此螺旋与组成这一元件的两个特异性的碱基对接触。鉴别子由 σ 区域 1、2 识别。

与启动子内的其他元件不同，UP 元件并不被 σ 因子所识别，而是被 α 亚基的羧基末端域（αCTD，图 12-7）识别。αCTD 通过一个可变形的衔接物与 α 亚基的氨基端域（αNTD）相连接。因而，虽然 αNTD 包埋在酶内部，αCTD 仍可以到达上游元件，即使该元件并不处于与－35 区域直接相邻的位置，而是位于上游更远处。

σ 亚基以这样一种方式定位在全酶结构内，实现对各种启动子元件的识别。所以，DNA 结合区域朝向酶体外部，而不是被包埋起来。而且，这些区域之间的间距与它们所识别的 DNA 元件之间的距离是一致的。当 σ 因子结合在全酶里时，σ 区域 2 和 4 分

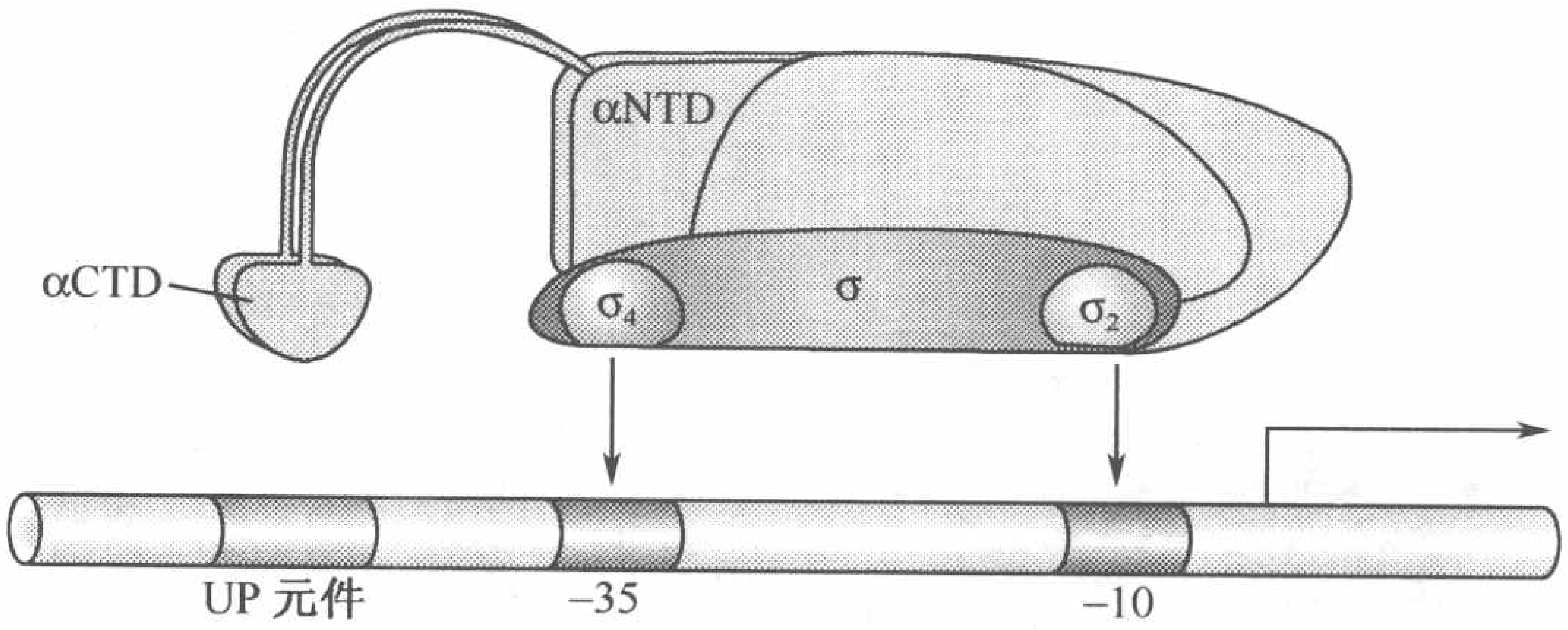


图 12-7 σ 和 α 亚基将 RNA 聚合酶核心酶募集到启动子上。 α 亚基的 C 端域（αCTD）识别 UP 元件（如果存在的话），而 σ 区域 2 和 4 分别识别－10 和－35区域（图 12-6）。在本图中，RNA 聚合酶的示意形式与在先前的图中所显示的有相当大的差别。这种形式对于显示与 DNA 和调节蛋白接触的面是非常有用的，我们在第 16 章考察细菌的转录调节时还将会在某些图中使用它。

开大约 75Å；这与一个典型的 $σ^{70}$ 启动子的－10 和－35 元件中心点之间的距离基本相同（图 12-7）。蛋白域之间的这一相当大的间距是由 σ 区域 2 和 4 之间的区域提供的，也就是区域 3，尤其是区域 3.2 也称为 σ3/4 连接区（图 12-4 和图 12-6）。

向开放式复合体的转变过程涉及 RNA 聚合酶和启动子 DNA 的结构变化

在闭合式复合体内 RNA 聚合酶与启动子 DNA 的结合使 DNA 仍处于双链状态。转录起始的下一个阶段要求聚合酶在开放式复合体中与启动子更紧密地结合在一起。从闭合式复合体到开放式复合体的转变过程涉及聚合酶的结构变化，以及 DNA 双链的打开，暴露出模板链和非模板链。相对于转录起始位点而言，这一“融化”发生在－11和＋3 之间的区域。

对于含 $σ^{70}$ 因子的细菌聚合酶，这一转变被称为异构化作用（isomerization）。它不需要从 ATP 水解获得能量，而是 DNA-酶复合体的构象自发地变成一种在能量上更具优势的形式。异构化作用本质上是不可逆的，一旦异构化完成，通常转录将会随后起始（虽然在某些情况下还会再发生调控）。相反，闭合式复合体的形成则是易于逆转的：聚合酶可以转变为开放式复合体，从启动子上脱离。

为描绘伴随异构化作用发生的结构变化，我们需要更为细致地了解全酶的结构。如前面所描述的，在这个爪状酶的钳子之间有一条通道穿过（图 12-2）。由 β 和 β′亚基的一些区域共同组成的活性位点，位于钳子基部的“活性中心裂隙”之内。

有 5 条通道通向酶的内部，如图 12-8 的开放式复合体的图片中所示。核苷三磷酸（NTP）摄取通道（图 12-8 的说明）允许核糖核苷酸进入活性中心。RNA 出口通道允许正在延长的 RNA 链在延伸合成的同时离开此酶。其他的 3 个通道允许 DNA 进出此酶，如下所述。

下游 DNA（也就是聚合酶前方将被转录的 DNA）以双链的形式经下游 DNA 通道（在两个钳子之间）进入活性中心裂隙。在活性中心裂隙内，DNA 双链从＋3 位置开始分离。非模板链通过非模板链（NT）通道离开活性中心裂隙并从聚合酶的表面经过，

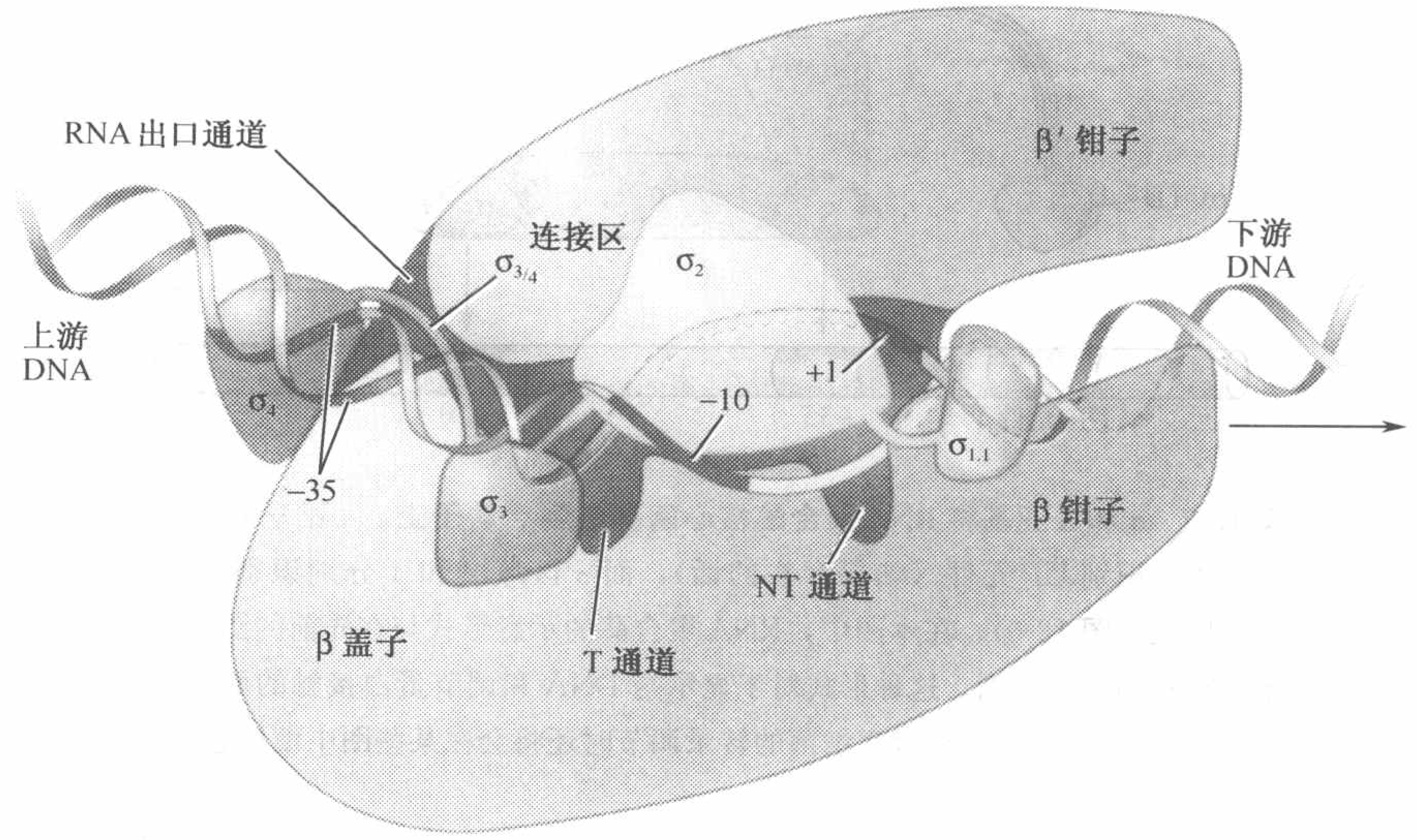


图 12-8 进出开放式复合体的通道。本图显示了两条 DNA 链的相对位置（模板链为灰色，非模板链为橘红色）、σ 因子的 4 个区域、启动子的－10 和－35 区域以及转录的起始位点（+1）。DNA 和 RNA 进出 RNA 聚合酶的通道也在图中显示。这里唯一没有显示的通道是核苷酸进入（N70－uptake）通道，核苷酸由此通道进入活性位点裂隙，结合到增长中的 RNA 链中。如图，该通道从页面深处进入位于 DNA+1 附近的活性位点。DNA 链从蛋白质下面经过的部分以虚线表示。σ 区域 3/4 连接区，也称 $\sigma_{3.2}$是 $\sigma_{3.1}$和 σ_4 之间的连接区域。（原图由 Richard Ebright 设计。）

模板链则沿着一条路线穿过活性中心裂隙，经模板链（T）通道离开。聚合酶后面的上游 DNA 在－11 位置处重新恢复双链状态。

从闭合式复合体向开放式复合体转变的异构化作用中，可以观察到聚合酶发生两个显著的结构变化。首先，位于聚合酶前部的钳子牢固地钳压在下游 DNA 上。第二，σ 的 N 端区域（区域 1.1）位置有至关重要的移动。在未结合 DNA 时，σ 区域 1.1 处于全酶的活性中心裂隙内部，阻断开放式复合体中模板 DNA 链经过的路线。在开放式复合体中，区域 1.1 移动了大约 50Å 的距离，出现在酶的外部，使 DNA 可以进到裂隙中（图 12-8）。σ 区域 1.1 处于高度的负电荷状态（恰似 DNA）。因而，在全酶中，区域 1.1 起着 DNA 的分子模拟物的作用。活性中心裂隙的空间呈高度正电荷状态，既可以被区域 1.1 占据，也可以被 DNA 占据。

转录由 RNA 聚合酶起始，不需要引物

第 8 章中谈到，DNA 聚合酶并不重新合成新的 DNA 链，也就是说，它只能延伸一条已经存在的多聚核苷酸链。因此，复制总是需要引物链。引物通常是一段较短的 RNA，与 DNA 模板链结合，形成一段较短的杂交双链区域；然后 DNA 聚合酶将核苷酸添加到引物的 3′端。

RNA 聚合酶能够在一条 DNA 模板上起始新的 RNA 链，并不需要引物。这一功能要求起始的核糖核苷酸被带到活性位点，并稳定地保持在模板上，同时下一个核苷三磷

酸以正确的几何形态呈递，使聚合的化学过程得以发生。这是相当困难的，因为 RNA 聚合酶的转录物大多以腺嘌呤核苷酸（A）开始，此核苷酸仅通过两个氢键与模板核苷酸（胸腺嘧啶核苷酸，T）结合，而胞嘧啶核苷酸（C）和鸟嘌呤核苷酸（G）之间有三个氢键。

因此，聚合酶必须与起始核苷酸或第二个核苷酸或两者都建立特异性的相互作用，严格控制其处于正确的方向，使其可以对引入的核苷三磷酸开始进行化学反应。这种对聚合酶与起始核苷酸之间特异性相互作用的要求可能解释了为什么大多数转录物都以同一种核苷酸开始。目前认为，此相互作用是由聚合酶全酶的不同部分提供的，包括 σ 的 3/4 链接处。与此一致的是，在用含有缺少此部分的 σ^{70} 的聚合酶进行的实验中，起始过程所需要的起始两个核苷酸中的一个或两个的浓度比平时要高得多。

转录起始阶段，RNA 聚合酶保持位置不变并将下游 DNA 拉向自己

如我们已经描述的，转录起始阶段，在逃离启动子、进入延长阶段、以及合成正确的转录物前，RNA 聚合酶合成并释放短于 10 个核苷酸的小 RNA 转录物（流产性合成）。很长时间为都不明白在最初的流产转录循环中，酶的活性位点是如何沿着 DNA 模板移动的。现在提出了三个通用模型（图 12-9）。

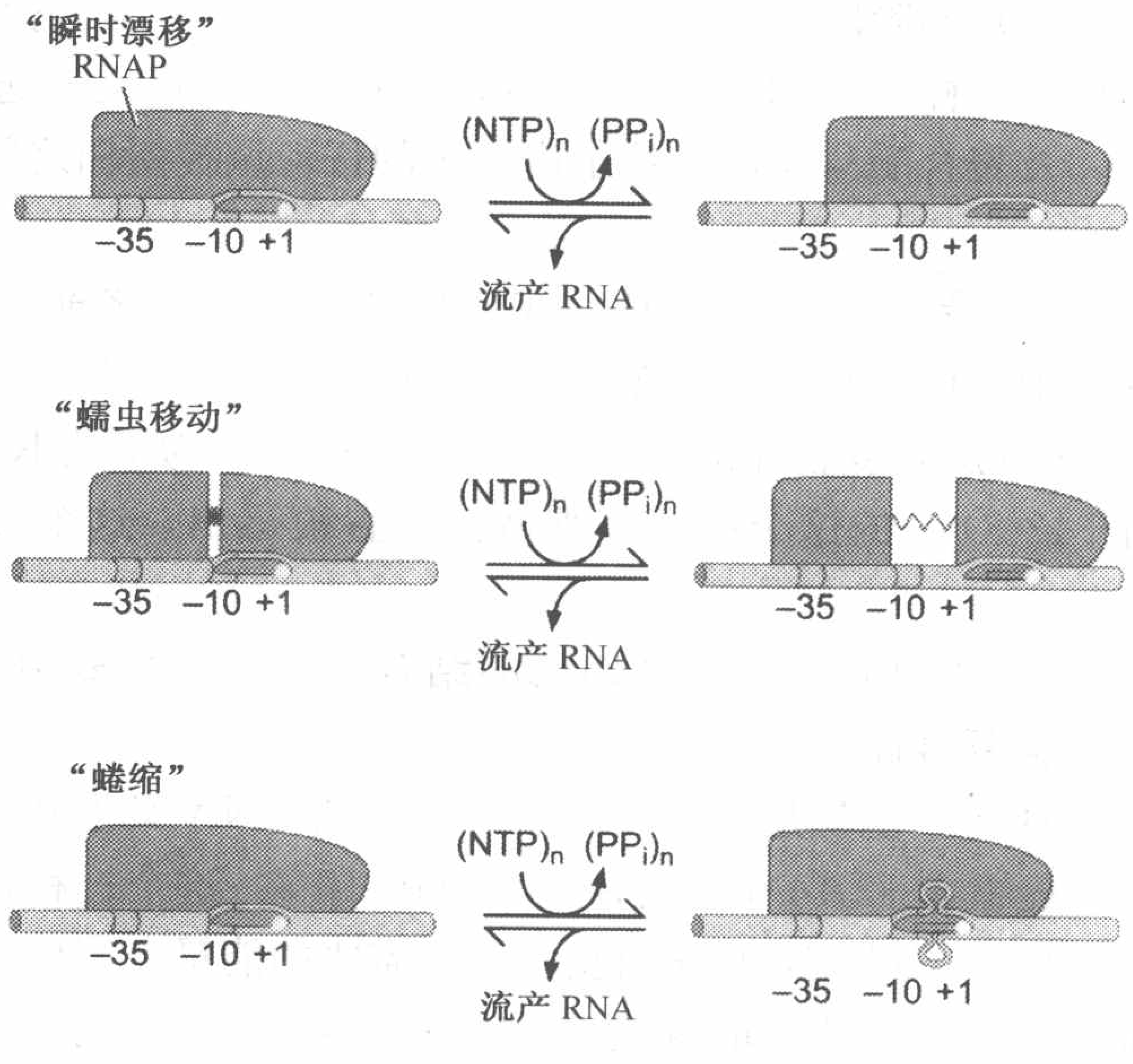


图 12-9 最初转录的机制。在最初的转录起始阶段，RNA 聚合酶的活性中心相对于 DNA 模板向前移动，合成短的转录物并流产，然后重复这个循环直到逃离启动子。图中表示了说明该过程的三个模型。根据第一个模型——瞬时漂移（上图）——聚合酶沿着 DNA 移动。在第二个模型中——蠕虫移动（中图）——酶的前部分沿着 DNA 移动，但由于酶内部有一个具伸缩性的区域，酶的后部分可以在启动子上保持固定。在第三个模型中——蜷缩（下图）——酶保持固定，把 DNA 拉过来。这些模型之间的不同点以及支持蜷缩模型的证据在正文中有描述。（来源：Kapanidis A. N. et al. 2006. Science 314：1144—1147，Fig. la. © AAAS 经许可修改）

1. 所谓的“瞬时漂移（transient excursion）”模型提了 RNA 聚合酶正向反向的瞬时迁移循环。聚合酶被认为离开启动子沿着 DNA 模板移动一小段，合成一小段转录物，然后流产，释放转录物，再回到其在启动子的原始位置。

2. “蠕虫移动（inchworming）”模型认为在聚合酶内有一个具伸缩性的元件，使得酶前方含有活性位点的部件向下游移动，合成一小段转录物，然后流产，再收缩回到其仍留在启动子上的酶主体。

3. “蜷缩（scrunching）”模型认为位于固定位置的、与启动子结合的、聚合酶下游的 DNA 被酶拉进来。于是在酶里堆积的 DNA 以单链泡形式逗留。

现在相信第三个模型（蜷缩）反映了真实的情况。这个结论基于一系列的发现，包括单分子分析实验，这些实验可以测量在最初的转录阶段聚合酶不同部分相互之间及相对于 DNA 模型的位置。结果显示，在最初的转录阶段，聚合酶在启动子上位置不变，解开下游 DNA 并把其拉进来。只有蜷缩模型与这些结果吻合。

启动子逃离涉及聚合酶-启动子相互作用及聚合酶核心-σ相互作用的破坏

我们已经看到，在最初的转录阶段，流产性起始过程发生，短的（9 个核苷酸或更短）转录物合成并释放。只有合成一条达到域值 10 或更多核苷酸的转录物时，聚合酶才能逃离启动子并进入延伸阶段。一旦达到这个长度，转录物不能停留在与 DNA 配对的区域，必须开始像线一样穿入 RNA 出口通道（图 12-8）。启动子逃离与打破聚合酶和启动子元件之间，及聚合酶与作用于特定启动子的任何调控蛋白之间的所有相互作用有关（第 16 章）。

目前还不清楚 RNA 聚合酶为什么必须在成功逃离启动子之前经历流产性转录起始。但是 σ 因子的一个区域参与其中，充当分子模拟物。这一次是 3/4 连接区，模拟的是 RNA。此区域位于开放式复合体的 RNA 出口通道中间，为了使 RNA 链可以合成超过 10 个核苷酸（图 12-8）σ 的这个区域必须从该位置被逐出，这个过程需要聚合酶尝试几次才能成功。

$\sigma_{3/4}$连接区的逐出可能说明了 σ 与延伸酶的结合比其与开发式复合体的结合更弱。事实上，它常常从延伸复合体上完全丢失。

一旦逃离，蜷缩逆转。也就是说，在蜷缩时解离的 DNA 又重新复合，伴随着转录泡从 22-24 个核苷酸萎缩成 12-14 个核苷酸大小（图 12-3）这个过程提供了聚合酶在逃离时打破聚合酶-启动子以及核心酶- σ 相互作用所需要的能量。因此，蜷缩是在转录起始中储存和转移能量的一种方法，能量在逃离时的释放可以使聚合酶与启动子分离及从核心酶驱逐 σ 因子。

在框 12-2 单亚基 RNA 聚合酶中，我们将看到这些简单的 RNA 聚合酶，尽管缺乏 σ 亚基，在从起始复合物向延伸复合物转变的过程中是如何经历结构上类似的转移的。

延伸聚合酶是一个边合成边校对 RNA 的向前行进的机器

DNA 通过延伸酶的形式与通过开放式复合体相似（图 12-8），双链 DNA 在两个钳子之间进入酶的前部。在催化裂隙的开口处，两条链分开并沿着不同的路径穿过酶，然后从各自的通道离开酶，并在延伸聚合酶（elongation polymerase）的后面重新形成双螺旋结构。核苷酸通过其固定的通道进入活性位点，并在模板 DNA 链的引导下添加到增长中的 RNA 链上。在任何一个特定时刻，增长中的 RNA 链中只有 8 个或 9 个核苷酸与 DNA 模板保持碱基互补状态；RNA 链的其余部分则剥落并从 RNA 出口通道离开酶。参见图 12-10 的延伸复合物示意图。

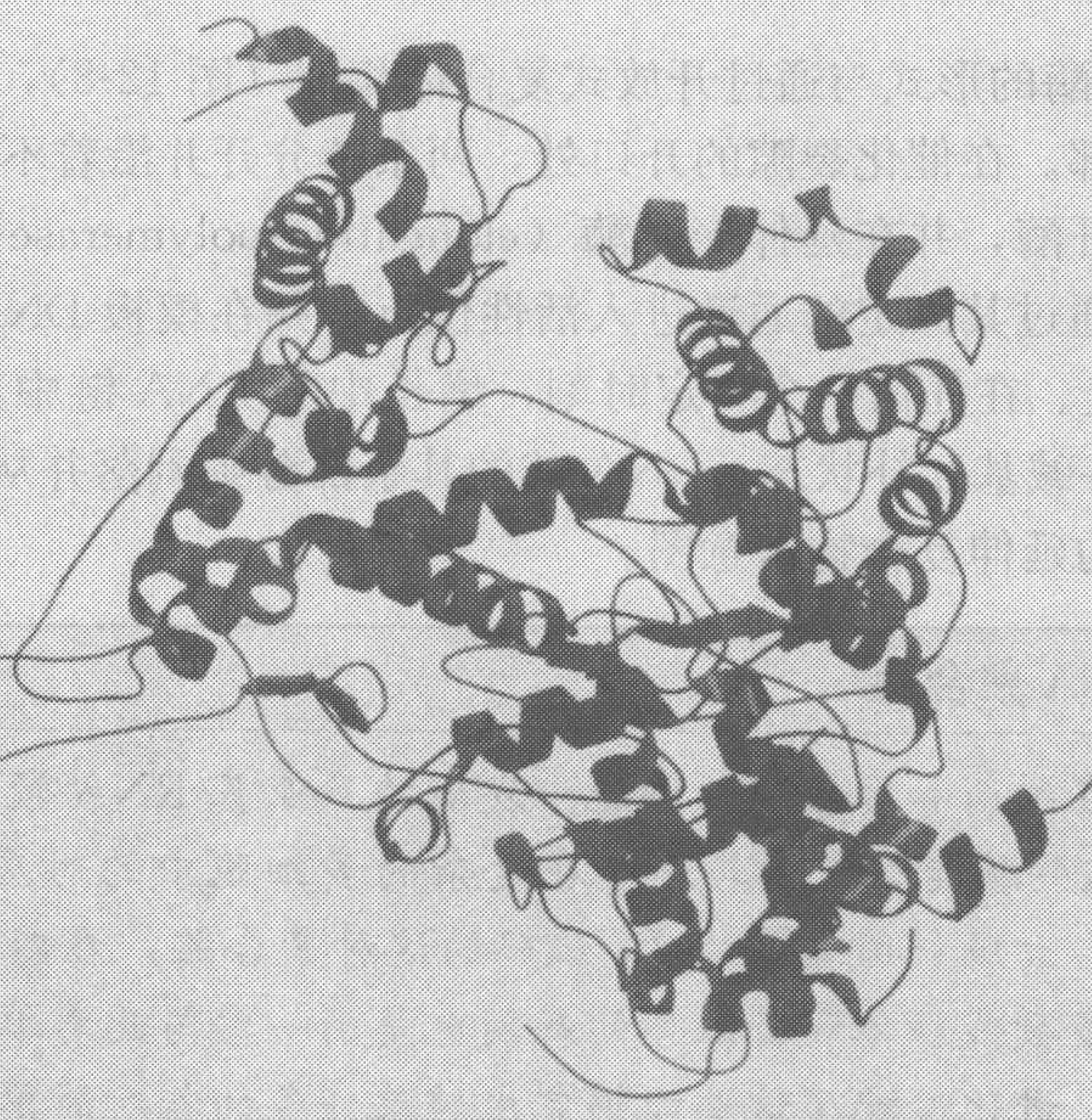
框 12-2　单亚基 RNA 聚合酶

正文中讨论了在细菌和真核生物细胞中发现的多亚基 RNA 聚合酶。但是亦有单亚基 RNA 聚合酶，这些酶能够完成与那些在复杂的多细胞中的对应物一样的基本反应。许多噬菌体，例如，大肠杆菌噬菌体 T7 编码的这种聚合酶，在感染时用以转录其大多数基因。类似地，多数线粒体和叶绿体的基因也是由一些与此类单亚基噬菌体酶紧密相关的聚合酶转录的。进化造就了这些能够完成转录任务的相对简单的酶，这是非常不平凡的。如我们在正文中强调过的，即使对大得多并且更为复杂的多亚基酶来说，转录都是一项令人印象深刻的成就。

T7 聚合酶是研究得最为广泛的单亚基聚合酶。它的分子质量为 100kDa，而细菌的核心酶为 400kDa（不包括 σ 因子），其结构在框 12-2 图 1 中显示。总体上它看起来与我们在第 8 章中描述过的 DNA 聚合酶的 Pol I 家族相似。因而，T7 的 RNA 聚合酶像是一只右手，具有手指、拇指和手掌，代表排布在中心裂隙周围的域，活性位点位于裂隙内。

虽然 T7 聚合酶在结构上与 RNA 聚合酶没有关联，反而更像 DNA 聚合酶，但它的确也具有与细胞的 RNA 聚合酶一样的特征。自从比较了 T7 聚合酶和细菌聚合酶在与各自的模板组成的复合体中的结构，这些特征已变得很明显。如我们在正文中所看到的，细菌的聚合酶具有各种各样的通道进出活性中心裂隙（图 12-8）。例如，一个通道允许核苷三磷酸进入活性中心，接触模板，在模板的影响下聚合到增长中的 RNA 链上。另一个通道则为增长中的 RNA 链提供了离开聚合酶的出口。类似的通道也可以在噬菌体聚合酶的结构中见到。

对细菌和 T7 聚合酶的起始和延伸复合体进行比较后发现一个明显的差异：即在细菌和 T7 中，一种类似的功能转变是通过不同种类的结构变化而实现的。我们在正文中曾特别指出，在细菌中，从起始到延伸的转变包括了 σ 因子一个区域的位置明显移动。这一移动打开了 RNA 出口通道，因此使长度超过 10 个核苷酸的转录物得以生成。T7 聚合酶中没有 σ 因子，但是在此单亚基酶体内的一个类似的结构变化介导了起始复合体向延伸复合体的转变，这一结构变化对于 RNA 出口通道的形成是必需的。



框 12-2 图 1 噬菌体 T7 的 RNA 聚合酶。(Jeruzalmi D. and Steitz T. A. 1998. EMBO J 17: 4101.) 图像使用 MolScript、BobScript 和 Raster 3D 等软件制作。

在延伸过程中，聚合酶在增长的 RNA 转录物上一次添加一个核苷酸。另外，泡的大小，即非双链 DNA 的长度，在延伸过程中保持不变：当一个核苷酸对在行进酶的前方分开，一个核苷酸对就在其后方形成。

除了所有这些功能外，RNA 聚合酶还执行着两项校对（proofreading）功能。第一个称为焦磷酸化编辑（pyrophosphorolytic editing)。在此状态下，聚合酶利用它的活性位点，在一个简单的逆向反应中，通过重新加入 PP_i，催化错误插入的核糖核苷酸的去除。然后聚合酶可以将另一个核糖核苷酸加入到增长中的 RNA 链的相应位置上。需要注意的是，以这种方式聚合酶既可以去除错误的碱基，也可以去除正确的碱基，但是它在不匹配的地方比在匹配的地方花的时间更长，所以常去除前者。在第二种称为水解编辑（hydrolytic editing）的校对机制中，聚合酶倒退一个或更多的核苷酸（图 12-10d）并切开 RNA 产物，去除掉错误的序列。

水解编辑是由一些 Gre 因子激发的。Gre 因子不仅能够增强水解编辑的功能，而且同时也作为延伸刺激因子。也就是说，它们确保聚合酶能够有效地延伸并帮助克服转录在较难转录的序列处“停滞不前”。这些功能的组合类似于真核生物的转录因子 TFIIS 对其 RNA 聚合酶的作用（见下文）。另一组蛋白质——Nus 蛋白，在延伸阶段加入聚合酶，并以目前还很不明确的方式促进延伸和终止的过程（见第 16 章延伸阶段中进行调控的例子)。

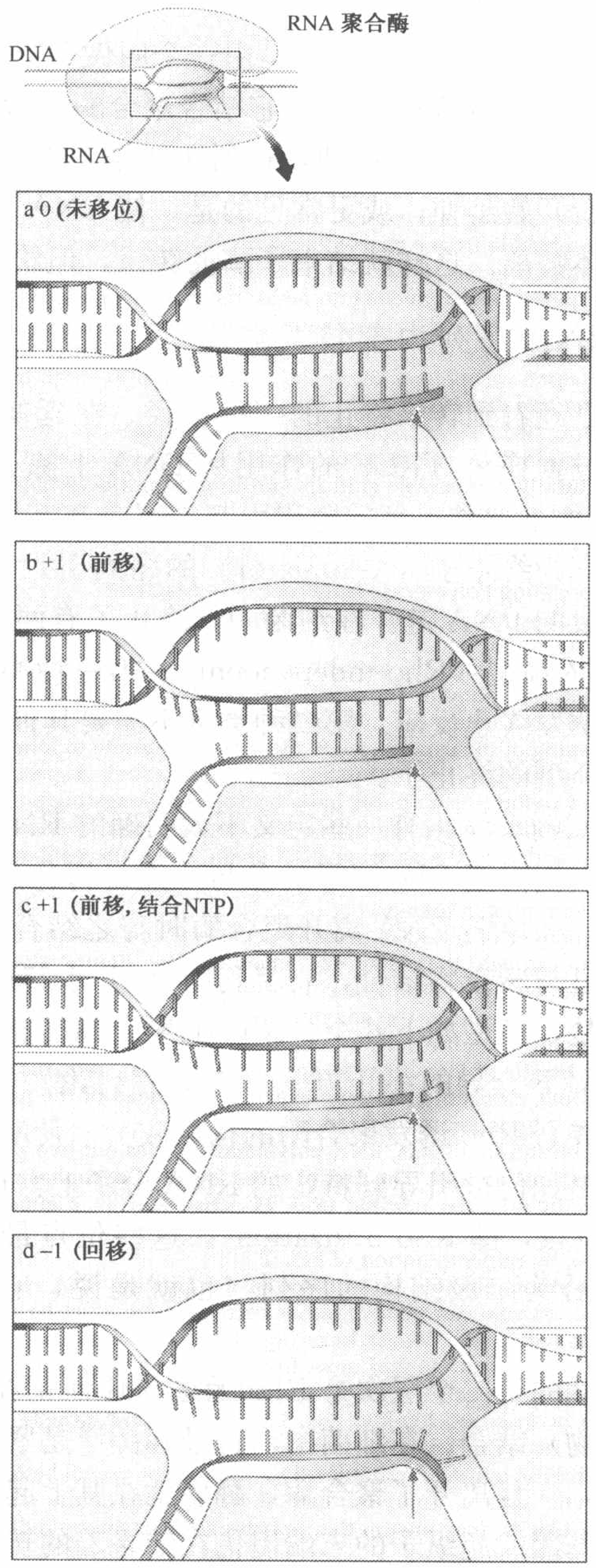


图 12-10 RNA 聚合酶延伸复合体内部的模板和转录物。图示为在转录过程的不同阶段 RNA 聚合酶内 RNA 及 DNA 模板相对位置的示意图。a. 未移位的聚合酶（0）显示 RNA 链与模板 DNA 链 9 个碱基的配对。b. 前移的聚合酶（+1）显示聚合酶已经向前方移动了一个碱基的情况。c. 结合 NTP 的前移聚合酶显示如图 b 中相同的 DNA 和 RNA 位置，带着一个进入的 NTP。d. 回移的聚合酶（−1）显示当聚合酶如在水解编辑时回移一个碱基的情况。红色箭头指向聚合酶中一个固定的位置，所有图中都一样。为了清晰起见，这里显示的聚合酶与图 12-8 方向不同，RNA 出口通道在下方。（来源：Figures based on images courtesy of Richard Ebright.）

RNA 聚合酶可能变得停滞不前，需要挪走

在有些情况下，一个延伸着的 RNA 聚合酶可能变得停滞不前并结束转录。停滞不前的一个常见理由是一条损坏的 DNA 链。如果被转录的基因是必不可少的，停滞不前的后果可能是灾难性的：没有产物会被停滞不前的聚合酶制造出来，并且对试图转录同一个基因的其他聚合酶造成路障。

为了处理这种情况，细胞具有挪走停滞不前的聚合酶并且同时招募修复酶（具体

地，内切酶 Uvr（A）BC）的机器。接下来的修复称为转录偶联的修复（第 9 章）。聚合酶的挪走及修复酶的招募都是由一个叫做 TRCF 的蛋白进行的。

TRCF 具有 ATase 活性。它结合聚合酶上游的双链 DNA，并且利用 ATPase 传动器（motor）沿着 DNA 移动，直到遇到停止的 RNA 聚合酶。碰撞把聚合酶往前推，或是使它重新开始延伸，或更常见的是，使 RNA 聚合酶、模板 DNA、和 RNA 转录物的三重复合体解离。这样就结束了酶的转录，但给另一个酶开辟了道路。

转录由 RNA 序列内部信号终止

我们已经看到转录终止的一种途径。在延伸过程中，当 RNA 聚合酶停滞不前时，它可以通过置换者 TRCF 的作用从 DNA 上除去（上面有讨论）。这种终止由损坏的 DNA 或其他不曾预料到的障碍引起。但是在基因的末端，转录终止是一个正常且重要的功能。称为终止子（terminator）的序列引发延伸聚合酶从 DNA 上脱离并且释放出它已经合成的 RNA 链。在细菌中，终止子有两种类型：Rho-依赖型（Rho-dependent）和 Rho-非依赖型（Rho-independent）。第一种类型，顾名思义，需要一个称为 Rho 的蛋白质来诱发终止反应。第二种类型不需要其他因子的参与就导致终止。我们将依次讨论这两种类型的终止子。

Rho-依赖型终止有一些定义不太明确的 RNA 元件，称为 rut 位点。如下面将要讨论到的，它们需要 Rho 因子的参与才能发挥作用。Rho 是一个具有 6 个相同亚基的环状蛋白质，在单链 RNA 离开聚合酶时与之结合（图 12-11）。此蛋白质还具有 ATPase 活性：一旦与转录物结合，它就利用 ATP 水解产生的能量来诱发终止。终止的准确机制还不清楚，模型如下几个：Rho 把聚合酶往相对于 DNA 和 RNA 的前方推，导致终止，与 TRCF 介导的终止方式（上面有讨论）类似；或 Rho 把 RNA 拉出聚合酶，导致终止；或者 Rho 诱发聚合酶的构象改变，使酶终止。

Rho 是如何被引导到特定的 RNA 分子上？首先，它结合的位点有某种特异性（所谓的 rut 位点，即 Rho utilization）。这些位点最理想的组成成分是几段约 40 个核苷酸、不折叠为二级结构的序列（也就是说，它们主要保持着单链状态），还富含 C 残基。

特异性的第二个层次是 Rho 无法与任何正在被翻译的转录物（也就是被核糖体结合的转录物）结合。在细菌中，转录和翻译是紧密箱连的——增长中的 RNA 转录物还在合成时，一旦离开了聚合酶，翻译就在其上起始了。因而，Rho 只会终止那些转录已超出一个基因或超纵子的末端但仍在转录着的转录物。

Rho-非依赖型终止子也称为固有终止子（intrinsic terminator），因为它们不需要其他因子就能发挥作用。Rho-非依赖型终止子由两个序列元件组成：一段短的反向重复序列（约 20 个核苷酸），其后是一段约 8 个 A：T 碱基对的序列（图 12-12）。这些元件要转录后才会影响聚合酶，也就是说，它们是以 RNA 而不是 DNA 起作用的。当聚合酶转录一段反向重复序列时，合成的 RNA 能够通过自身碱基配对形成一个茎-环结构（stem-loop）（常称为“发夹”，第 6 章）。发夹的形成破坏了延伸复合体，从而引起终止。与 Rho-依赖型终止一样，其作用机制还不清楚。现行的模型与那些 Rho 依赖型的

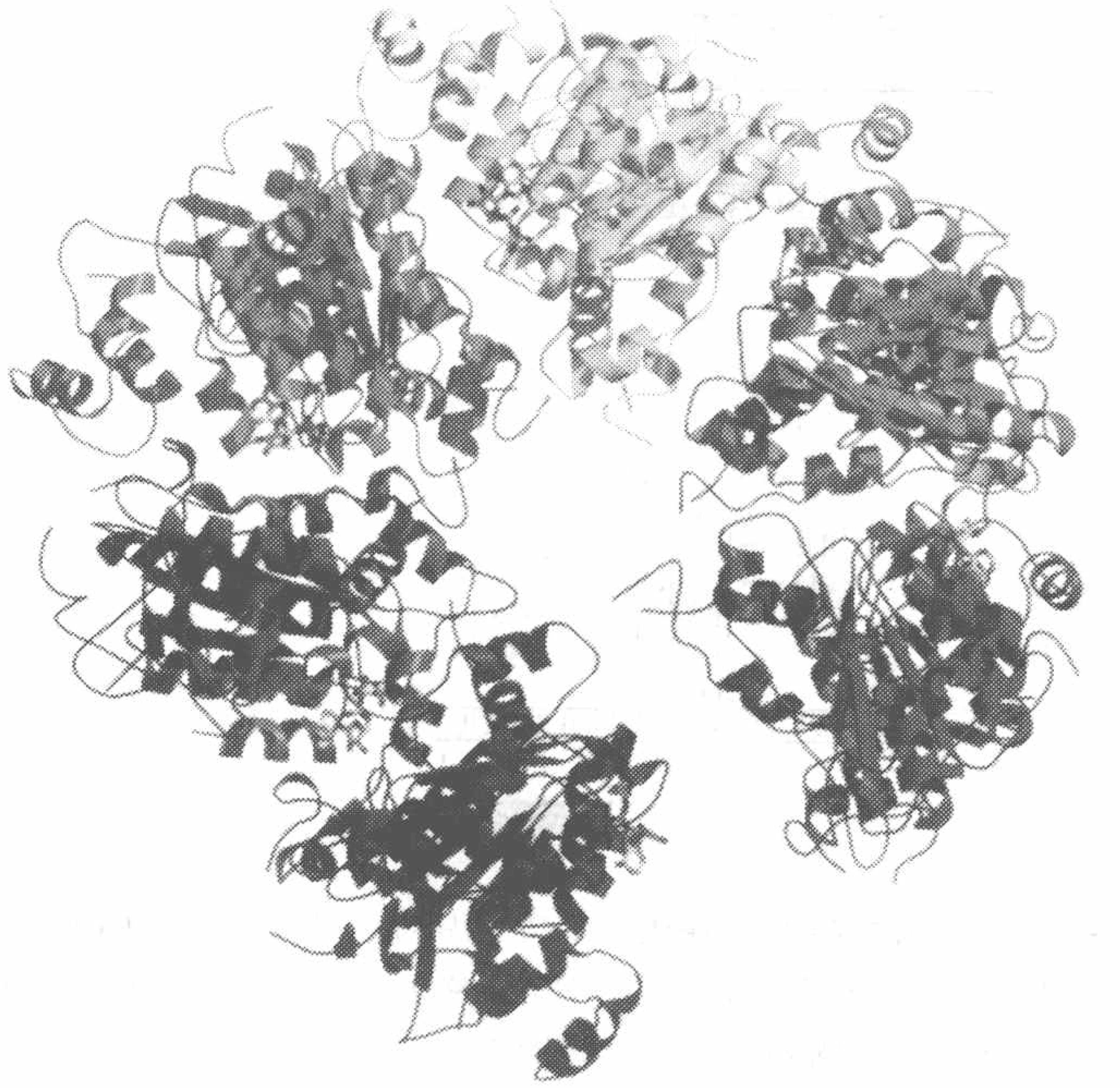


图 12-11 ρ转录终止因子。图为 Rho 终止因子的晶体结构的俯视图。它由一个 Rho 蛋白的六聚体组成，每一个单体以一种不同的颜色表示。六个单体形成一个开放的环，这个环并不是平面的，第六个亚基要比第一个更深入页面所在的平面。两个亚基之间的缝隙为 12Å，螺旋倾斜度为 45Å。据认为 Rho 所作用的 RNA 转录物（未显示）是沿着每个亚基的底部结合的，然后从环的中央穿过。(Skordalakes E. and Berger J. M. 2003. Cell 114：135.）图像使用 MolScript、BobScript 和 Raster 3D 等软件制作。

差不多。也就是说，发夹或是把聚合酶往相对于 DNA 和 RNA 的前方推，或是把转录物从聚合酶上硬扯下来，或是诱发聚合酶的构象改变而导致终止。

如我们以前所描述过的，发夹结构只有在后面跟随着一段 A：U 碱基对时才能有效地发挥终止子的作用。这是因为，在这种环境中，当发夹结构形成时，增长中的 RNA 链会仅通过 A：U 碱基对固定在模板上的活性位点。由于 A：U 碱基对是所有碱基对中最弱的（甚至比 A：T 碱基对还要弱），所以更容易被茎-环结构对转录聚合酶的作用所打断，RNA 也因而更容易脱离下来（图 12-13）。

真核生物的转录

如我们已经讨论过的，真核生物中的转录是由与细菌中发现的 RNA 聚合酶密切相关的一类聚合酶完成的。这几乎不会让人感到惊讶：转录过程的本身在原核生物与真核生物中是相同的。然而，每种情况中所用的“机器”存在差别。有一点我们已经看到了：真核生物有三个不同的聚合酶（Pol Ⅰ、pol Ⅱ和 pol Ⅲ），而细菌只有一个。还有，细菌只需要

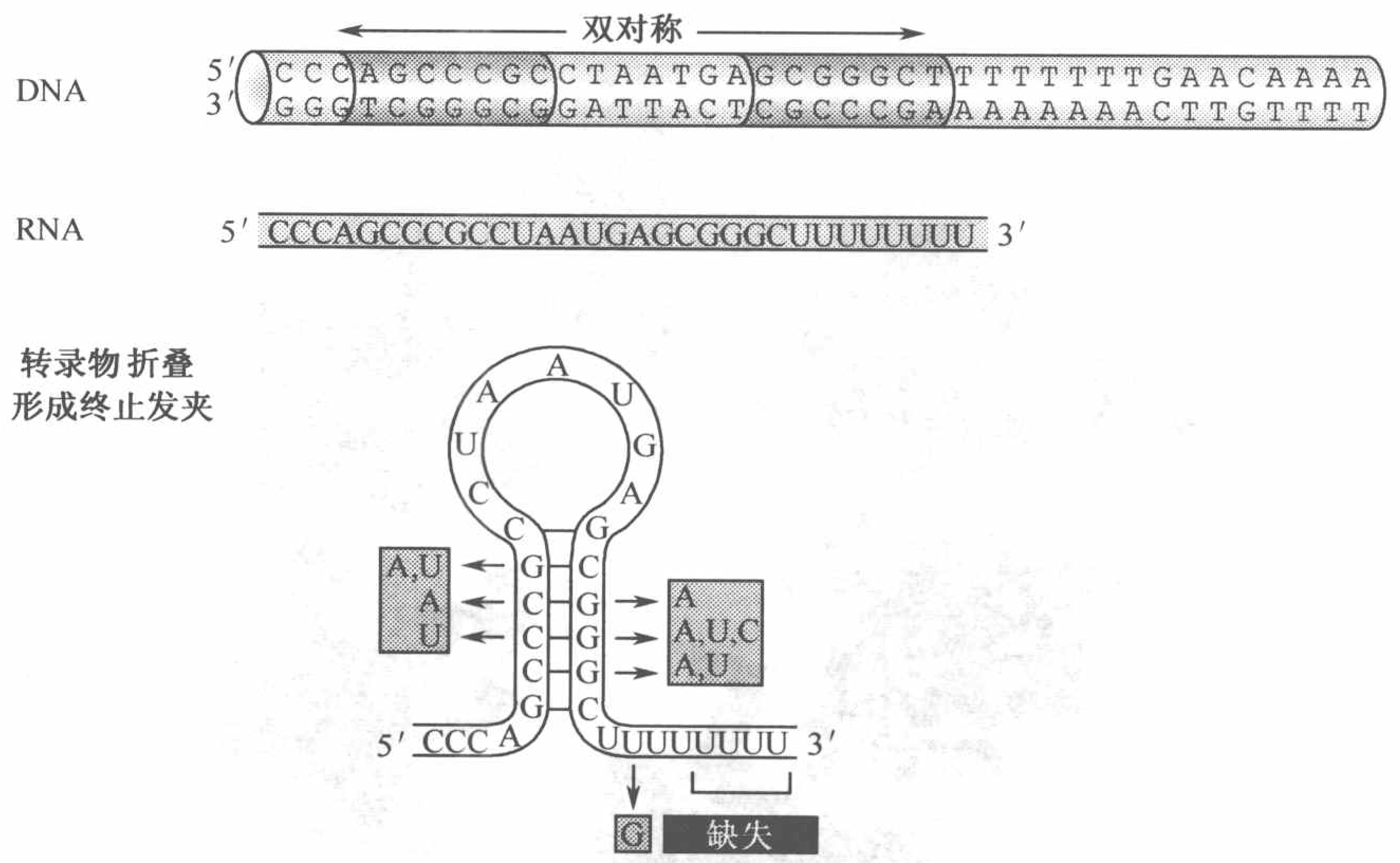


图 12-12 一个 rho—非依赖型终止子的序列。顶上是终止子在 DNA 中的序列，其下显示的是 RNA 的序列，底部是终止子发夹的结构。所讨论的终止子来自于第 16 章中讨论的色氨酸衰减子（trp attenuator）。方框中显示的是分离到的破坏终止子序列的突变。（来源：Adapted from Yanofsky C. 1981. Nature 289：751-758）

一个额外的起始因子（σ），而真核生物中有效的和启动子特异的起始则需要几个起始因子。这些起始因子称为通用转录因子（general transcription factor，GTF）。

在体外，通用转录因子和 Pol Ⅱ构成从一个 DNA 模板（不含组蛋白）上起始转录所需要的全部材料。然而，如我们在第 7 章中所见，在体内，真核细胞内的 DNA 模板是整合在核小体内的。在这种环境下，单靠通用转录因子就不足以结合启动子序列并启动有意义的表达，而是需要额外的因子参与，包括 DNA 结合调节蛋白、所谓的“中介蛋白复合体”（mediator complex）和染色质修饰酶。

我们将先讨论在体外 Pol Ⅱ和通用转录因子组合在一个启动子上起始转录的基本机制，然后再讨论在体内启动转录所需的额外成分的作用。

RNA 聚合酶Ⅱ的核心启动子由 4 个不同序列的元件组合而成

真核细胞的核心启动子（core promoter）是指在体外检测时，Pol Ⅱ机器精确地起始转录所需要的最少的一组序列元件。代表性的核心启动子的长度大约有 40 个核苷酸，从转录起始位点向上游或下游延伸。图 12-14 显示了在 Pol Ⅱ核心启动子中发现的几个元件相对于转录起始位点的位置。它们分别是 TFIIB 识别元件（TFIIB recognition element，BRE）、TATA 元件（或盒子）、起始子（initiator，Inr）和下游启动子元件（downstream promoter element，有 DCE，MTE，DPE）。通常，一个启动子只含有这些元件之中的几个。比如，启动子通常含有 TATA 元件或 DPE，而不是两者。含 TATA 的启动子经常也会含有一个 DCE。Inr 是最常用的元件，与 TATA 和 DPE 都能共同存在。每一个元件的共有序列和与其结合的通用转录因子在图中也有显示，我们将在下面的部分对这些特征进行更为详细的描述。

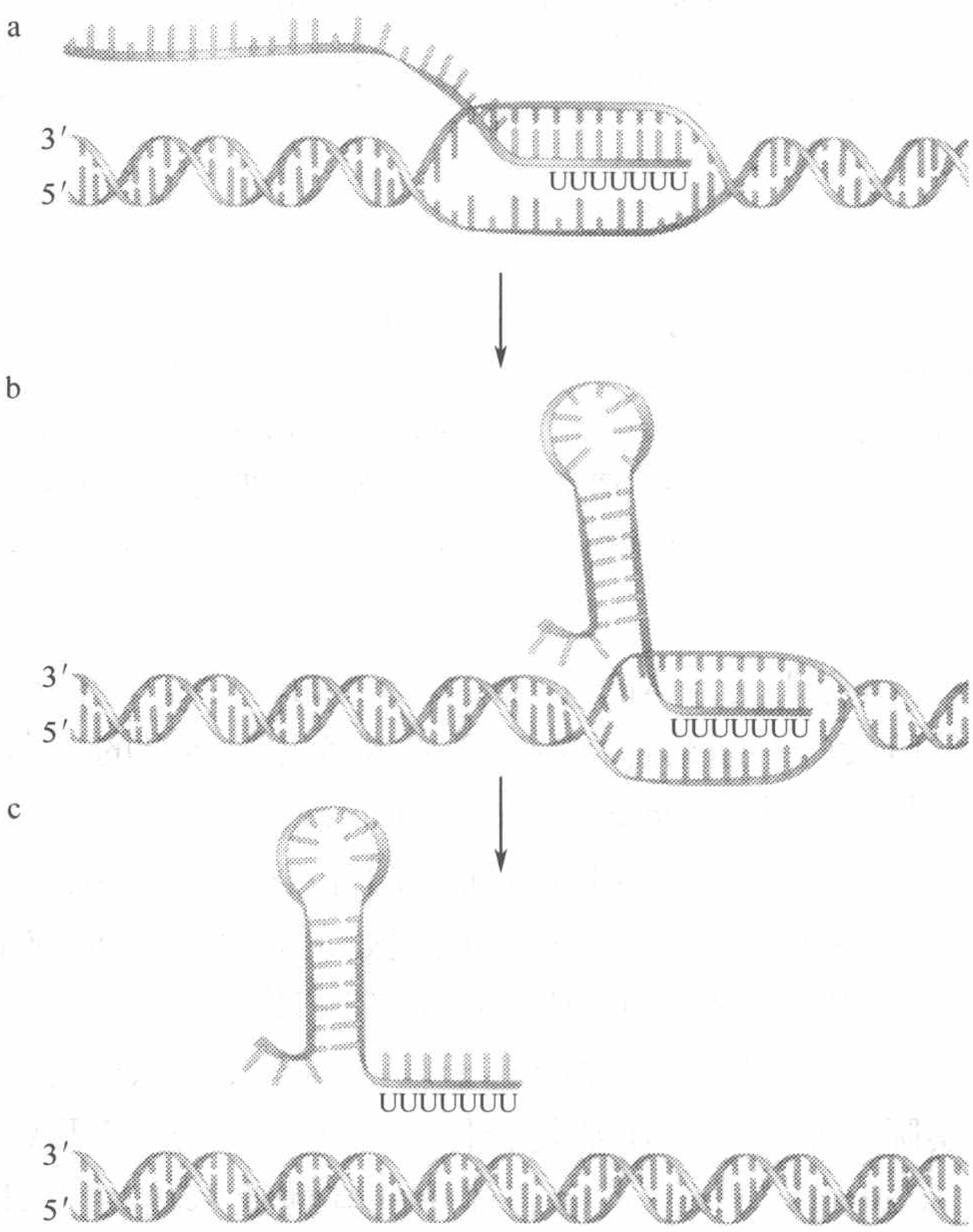


图 12-13 转录终止。图为 rho-非依赖型终止子如何工作的模型。a. RNA（图 12-10）中的发夹结构在该区域被聚合酶转录后立刻形成（聚合酶未在此显示）；b. 该 RNA 结构在聚合酶转录 DNA 下游的 AT 富集区域时将其破坏；c. 究竟发夹是如何破坏正在转录的聚合物还不太清楚（文中有不同的模型）发夹结构以及 RNA 中一段 U 与模板中 A 之间的弱相互作用协同显然使转录物的释放变得容易（来源：Adapted from Platt T. 1981，Cell 24：10-23.）

在核心启动子之外（而且通常是在其上游）存在一些在体内进行有效转录所需的其他序列元件。这些元件共同组成调节序列（regulatory sequence），并可以划分为几类，反映出它们的位置、所讨论的生物体以及它们的功能。这些元件包括：启动子最近元件（promoter proximal element），上游激活物序列（upstream activator sequence，UAS），增强子（enhancer），以及一系列称为沉默子（silencer）、边界元件（boundary element）和绝缘子（insulator）的抑制元件。所有这些 DNA 元件都与调控蛋白（激活因子或抑制因子）结合。调控蛋白可以帮助或阻碍从核心启动子开始的转录，是第 17 章的主题。其中有些调控序列可以位于距它们所作用的核心启动子几万甚至几十万碱基之处。

RNA 聚合酶Ⅱ在启动子上和通用转录因子形成前起始复合体

通用转录因子共同执行 σ 因子在细菌转录中所执行的功能其中有些因子与 σ 的不同区域具有微弱的同源性。因而，通用转录因子帮助聚合酶结合到启动子上并解开 DNA

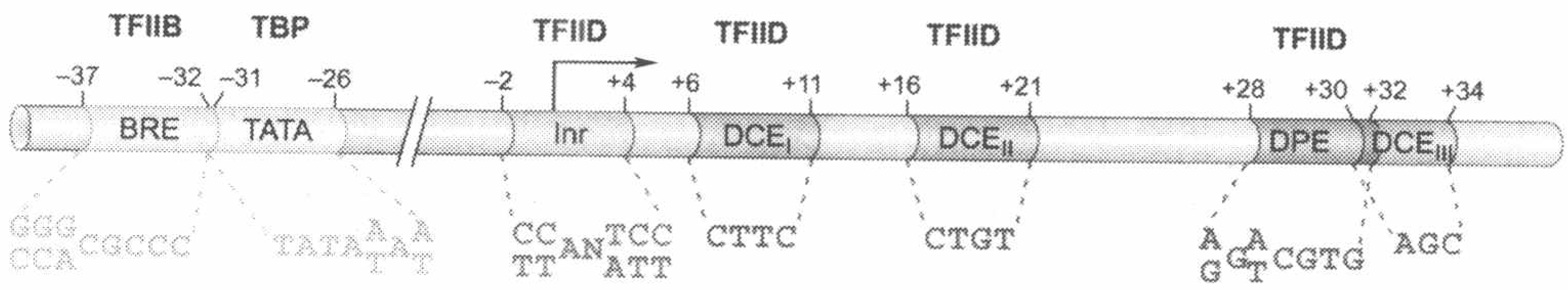


图 12-14 Pol Ⅱ 核心启动子。图中显示了不同 DNA 元件相对于转录起始位点(以 DNA 上方的箭头指示)的位置。正文中所描述的这些元件如下:BRE(TFIIB recognition element);TATA(TATA Box);Inr(initiator element);DPE(downstream promoter element);和 DCE(downstream core element)。另一个在正文中提及的元件,MTE(motif ten element),未在本图显示,但就位于紧邻 DPE 的上游。还显示了每个元件的共有序列(元件的确定方法与细菌启动子元件相同,框12-1),以及(上方)识别每一个元件的通用转录因子的名称。

(类似于细菌转录中由闭合式复合体向开放式复合体的转变)。它们也帮助聚合酶从启动子上逃离并开始延伸阶段。一起结合在启动子上准备开始转录的一套完整的通用转录因子和聚合酶称为前起始复合体(pre-initiation complex)。

如上所述(以及图 12-14),许多 Pol Ⅱ启动子包含一个所谓的 TATA 元件(大约在转录起始位点上游 30 个碱基对处)。这是前起始复合体开始形成的位置。TATA 元件是由叫做 TFIID 的通用转录因子所识别的。[“TFII”的名称是指 Pol Ⅱ的转录因子(transcription factor for Pol Ⅱ),单独的因子以 A、B 等加以区分。] 像很多通用转录因子一样,TFIID 实际上是一个多亚基复合体。TFIID 中与 TATA 序列结合的成分称为 TBP(TATA binding protein)。此复合体中的其他亚基称为 TAF,即 TBP 关联因子(TBP-associated factor)。某些 TAF 识别其他启动子核心元件,如 Inr、DPE 和 DCE,虽然最强的结合是在 TBP 和 TATA 之间。因此,TFIID 是启动子识别和建立前起始复合体的必不可少的因子。

TBP 一旦结合到 DNA 上,就使 TATA 序列极大地扭曲变形(我们不久将对这一点进行更为详细的讨论)。形成的 TBP-DNA 复合体提供了一个平台,把其他通用转录因子和聚合酶本身募集到启动子上。在体外,这些蛋白质按照下列顺序在启动子处组装(图 12-15):TFIIA、TFIIB、TFIIF 与聚合酶一起,然后是 TFIIE 和 TFIIH。包含这些成分的前起始复合体形成后,启动子就开始解旋。与细菌中的情形相反,真核细胞中的启动子解旋需要 ATP 的水解并由 TFIIH 介导,是该因子的类解旋酶活性刺激了启动子 DNA 的解开。

启动子逃离需要聚合酶“尾巴”的磷酸化

与细菌中的过程类似,在聚合酶从启动子逃离并进入延伸阶段之前,有一个流产性起始时期。回忆一下,在流产性起始过程中,聚合酶合成一连串短的转录物。在真核细胞中,启动子逃离包含在细菌中所没有见到的两个步骤:一个是 ATP 水解(除了前面 DNA 解离所需要的 ATP 水解之外),另一个是聚合酶的磷酸化,对此我们即将进行描述。

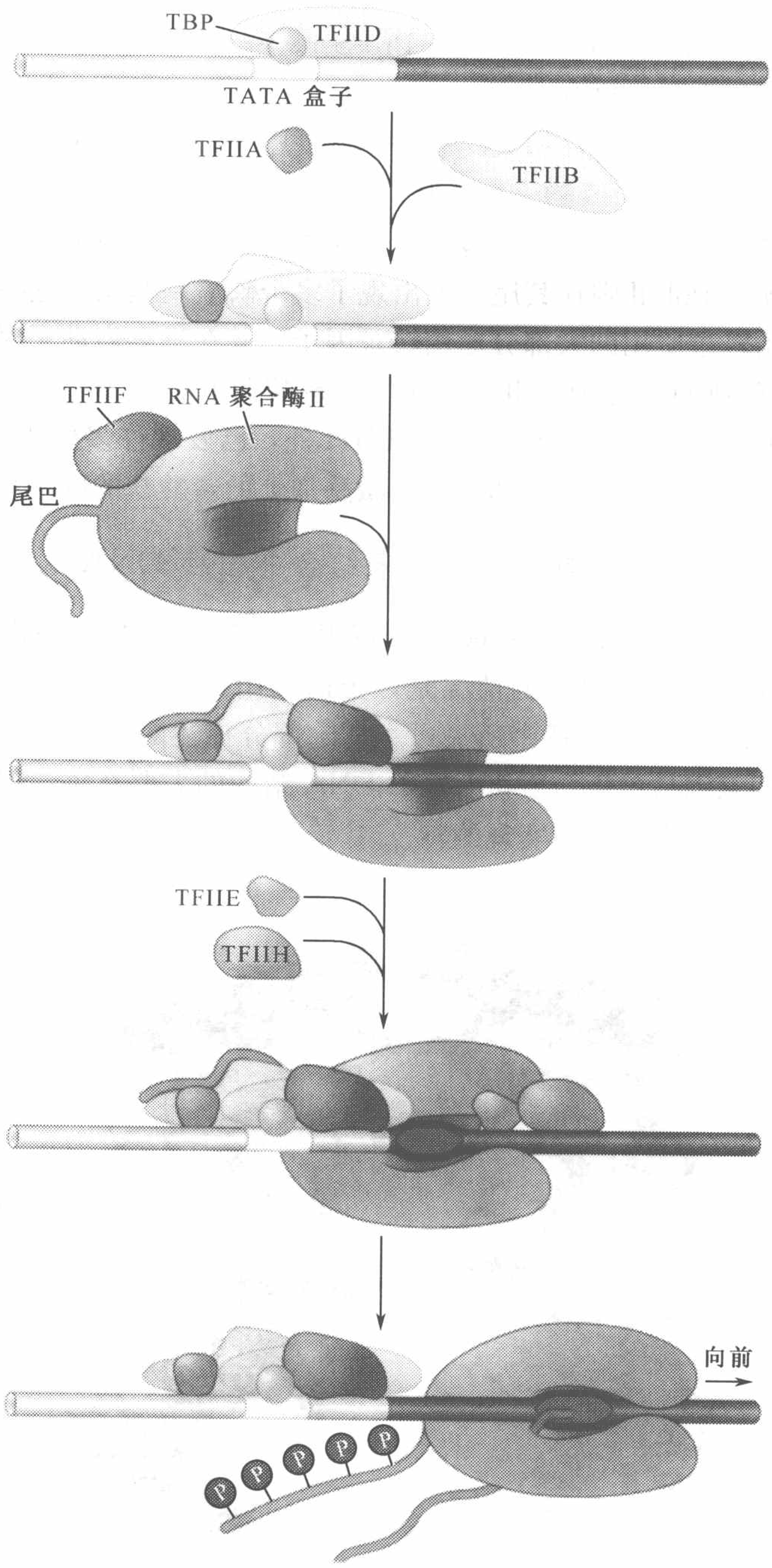


图 12-15　由 RNA 聚合酶Ⅱ进行的转录起始。这里显示了按步骤进行的 Pol Ⅱ前起始复合体的组装过程，正文对此进行了详细描述。一旦在启动子处完成组装，添加 RNA 合成所需的核苷酸前体以及聚合酶“尾巴”内部的丝氨酸残基发生磷酸化，Pol Ⅱ就会离开前起始复合体。此尾巴包含多个重复的七肽序列：Tyr-Ser-Pro-Thr-Ser-Pro-Ser（见图 12-19）。

Pol Ⅱ的大亚基有一个 C 端域（CTD），称为“尾巴”（图 12-15）。CTD 包含一连串重复的七肽序列：Tyr-Ser-Pro-Thr-Ser-Pro-Ser。酵母 Pol Ⅱ的 CTD 中有 27 个这样的重复序列，秀丽新小肝线虫有 32 个，果蝇有 45 个，人类有 52 个。事实上，重复序列的数字看来与基因组的复杂性相关。每一个重复序列都包含特定蛋白激酶的磷酸化位点，其中一个激酶就是 TFIIH 的一个亚基。

被募集到启动子上的 Pol Ⅱ最初包含一个在很大程度上未被磷酸化的尾巴，但是在延伸复合体里发现的 Pol Ⅱ则在其尾巴上出现了多个磷酰基基团。这些磷酸盐的添加帮助聚合酶摆脱起始转录所用的大部分通用转录因子，并在逃离启动子时将它们丢下。

如我们将要看到的，对 Pol Ⅱ的 CTD 的磷酸化状态的调控也控制着后面的步骤——延伸甚至 RNA 加工。实际上，除 TFIIH 外，已发现许多作用于 CTD 的其他蛋白激酶，并且还发现了一个去除由那些蛋白激酶加上的磷酸盐的磷酸酯酶。

TBP 用一个插入双螺旋小沟的β折叠来结合并扭曲 DNA

TBP 用一个广泛的β折叠（β sheet）区域识别 TATA 元件的小沟（minor groove，图 12-16），这是不常见的。典型的情况是，蛋白质用插入 DNA 大沟的α螺旋来识别 DNA，如我们在第 5 章和第 6 章中所见，本章前部分描述的σ因子也是如此。TBP 的非传统识别机制的原因与该蛋白质扭曲局部 DNA 结构的需要有关。但是这一识别模式引出了一个问题：特异性是如何实现的？

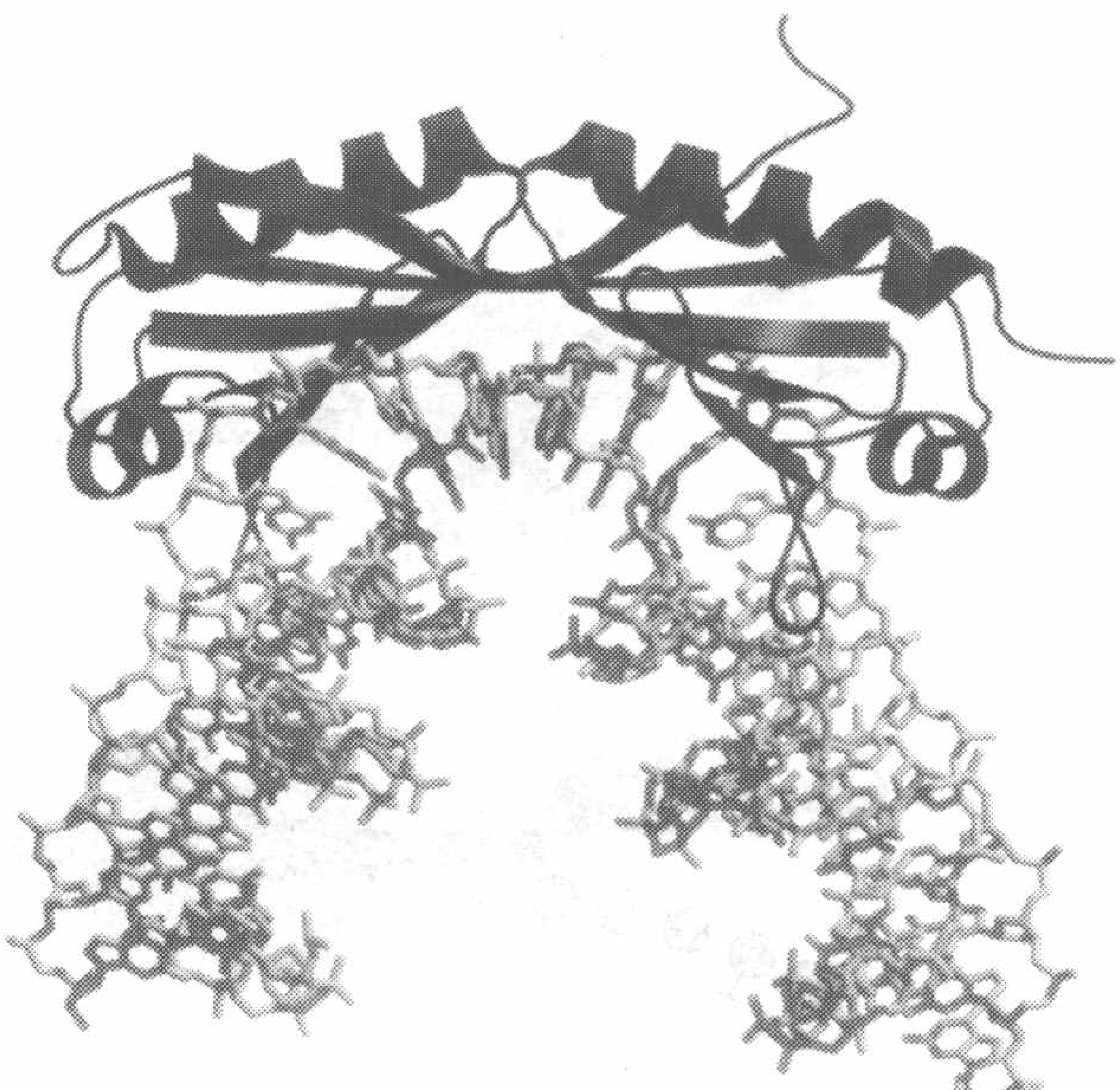


图 12-16　TBP-DNA 复合体。TBP（以紫色显示）与在许多 Pol Ⅱ基因的起始处发现的 DNA 的 TATA 序列（以灰色显示）构成复合体，这一相互作用的细节在正文中进行了描述。（Nikolov D. B. et al. 1995. Nature 377：119.）图像使用 MolScript、BobScript 和 Raster 3D 等软件制作。图像两侧延伸的 DNA 模型由 Leemor Joshua-Tor 建立。

我们在第 6 章中已经看到，与 DNA 的大沟相比，小沟中能够区分碱基对的化学信息较少。相反，TBP 依赖 TATA 序列能够经受特殊结构扭曲的能力而对其进行选择，如下面即将要描述的。

当与 DNA 结合时，TBP 会使小沟扩展为一个几乎是平面的构象；它还以大约 80°的角度使 DNA 弯曲。TBP 与 DNA 的相互作用仅包括该蛋白质和小沟中碱基对边缘之间的有限的一些氢键。相反，特异性主要由插入到识别序列两端的碱基对之间并驱使 DNA 发生强烈弯曲的两对苯基丙氨酸（phenylalanine）侧链提供。

因而，A∶T 碱基对是更有利的，因为它们易于扭曲而使小沟开始开放。β 折叠中的磷酸盐骨架和基本残基之间也存在着广泛的相互作用，从而增加了相互作用的总体结合能量。

其他通用转录因子在起始中也有特殊作用

我们并不详细知道所有其他通用转录因子的功能。如我们已经指出的，这些因子中的一些实际上是由两个或更多亚基组成的复合体（表 12-2）。下面我们来评论几个结构和功能的特征。

TAF TBP 与大约 10 个 TAF 结合。两个 TAF 在启动子处结合 DNA 元件，如起始子元件（Inr）和下游启动子元件（DPE）（图 12-14）。几个 TAF 与组蛋白有结构同源性，并且，尽管没有证据，已有人提出它们可能以相似的方式与 DNA 结合。例如，果蝇（*Drosophila*）的 TAF42 和 TAF62 已被表明形成一个类似于H3 · H4四聚体（第 7 章）的结构。这些类组蛋白 TAF 不仅在 TFIID 复合体中发现，而且与一些组蛋白修饰酶结合，如酵母的 SAGA 复合体（表 7-7）。

表 12-2 RNA 聚合酶Ⅱ的通用转录因子

GTF	亚基数目
TBP	1
TFIIA	2
TFIIB	1
TFIIE	2
TFIIF	3
TFIIH	9
TAF	11

注：表中所列数目为酵母的情况，但其他真核细胞，包括人类的，也很近似。然而还是有些不同点，如人类 TFIIF 只有 2 个亚基，TFIIA 有 3 个亚基。

另一个 TAF 似乎调控 TBP 与 DNA 的结合。它利用一个与 TBP 的 DNA 结合表面结合的抑制性盖子来完成这一任务——这是另外一个分子模拟的例子。TBP 要与 TATA 结合，这个盖子就必须被移去。

TFIIB 此蛋白质为一条多肽单链，在 TBP 之后进入前起始复合体（图 12-14）。TFIIB-TBP-DNA 三元复合体的晶体结构显示出特殊的 TFIIB-TBP 和 TFIIB-DNA 联系（图 12-17），其中包括与 TATA 元件上游的大沟（至 BRE，图 12-14）和下游的小沟之间的碱基特异性相互作用。TFIIB 与 TBP-TATA 复合体的非对称性结合造成了前起始复合体其余部分组装的非对称性以及由此引起的单向转录。TFIIB 在前起始复合体中也与 Pol II 接触。因而，此蛋白质似乎是连接结合了 TATA 的 TBP 与聚合酶之间的桥梁。结构研究显示 TFIIB 的部分区域以与细菌中 $\sigma_{3/4}$ 连接区相似的方式插入 Pol Ⅱ的

RNA 出口通道和活性中心裂隙。



图 12-17 TFIIB-TBP-启动子复合体。此结构显示了结合在 TATA 序列上的 TBP 蛋白，正如我们在前面的图中所看到的。这里，通用转录因子 TFIIB（以绿松色表示）已经加上了。这个三重复合体形成了一个平台，在前起始复合体组装过程中，其他通用转录因子和 Pol Ⅱ本身被募集到此平台上（Nikolov D. B.，et al. 1995. Nature 377：119.）图像使用 MolScript、BobScript 和 Raster 3D 等软件制作，图像两侧延伸的 DNA 模型由 Leemor Joshua-Tor 建立。

TFIIF 这个双亚基（人类）因子与 Pol Ⅱ结合并与 Pol Ⅱ（以及其他因子）一起被募集到启动子上。Pol Ⅱ与 TFIIF 的结合起着稳定 DNA-TBP-TFIIB 复合体的作用，并且是 TFIIE 和 TFIIH 被募集到前起始复合体上的前提条件（图 12-15）在酵母中，这个因子包括第三个亚基（表 12-2），但第三个亚基的功能未知。

TFIIE 和 TFIIH TFIIE 与 TFIIF 相似，由两个亚基构成，它是下一个结合到前起始复合体上的，并参与 TFIIH 的募集以及调控。TFIIH 控制着依赖 ATP 的从前起始复合体向开放式复合体的转变。它也是通用转录因子中最大和最复杂的——它有 10 个亚基，而且分子质量与聚合酶本身相当。TFIIH 内有两个亚基具有 ATP 酶的功能，还有一个亚基是蛋白激酶，参与启动子的解旋和逃离，如上面所描述过的。与其他因子一起，ATP 酶亚基还参与核苷酸错误匹配的修复（第 9 章）。

体内的转录起始需要其他蛋白质参与，包括中介蛋白复合体

到现在为止，我们已经描述了在体外 Pol Ⅱ从一条裸露的 DNA 模板起始转录所需的条件。但是我们已经指出，细胞内高水平的、受调控的转录还额外需要转录调控蛋白、中介蛋白复合体和核小体修饰酶（其自身常常是大的蛋白质复合体的组成部分）（图 12-18）。各修饰复合体的特征列在表 7-8 中。

这些额外需求的原因之一是 DNA 模板在细胞内包装在染色质内，正如我们在第 7

章中讨论过的。这一条件使聚合酶及其相关因子与启动子的结合变得复杂。称为激活因子的转录调控蛋白帮助将聚合酶募集到启动子上，使其稳定地结合在那里。这一募集过程是通过DNA结合的激活因子、染色质修饰及重塑因子、及转录机器组成部分之间的相互作用介导的。其中一个相互作用是与中介复合体（因此而得名）。中介蛋白通过一个表面与大聚合酶亚基的CTD“尾巴”结合，而将其他表面用于与DNA结合的激活因子之间的相互作用。这解释了在细胞内成功地进行显著的转录需要中介蛋白的原因。

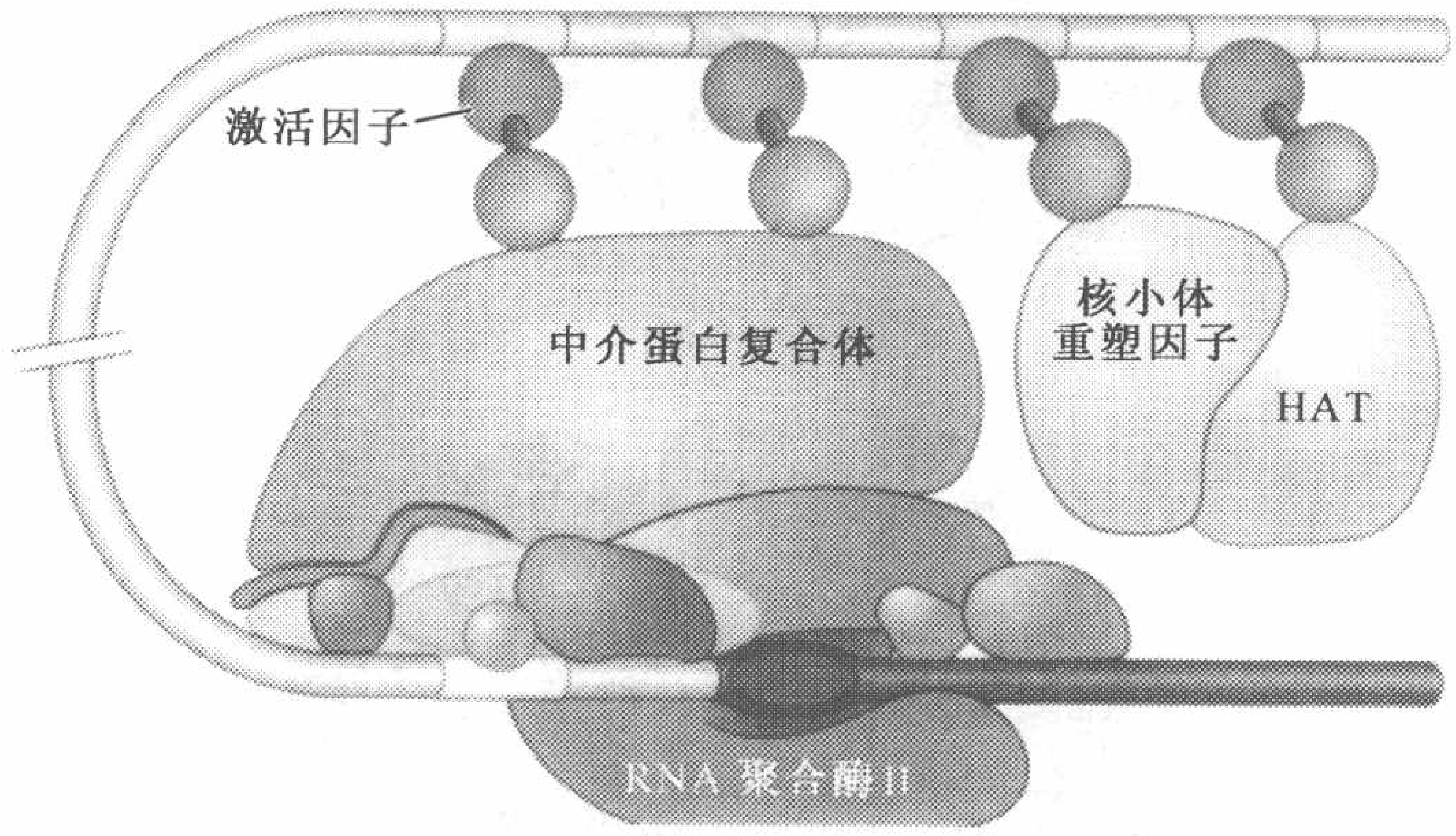


图 12-18 在中介蛋白、核小体修饰因子和重塑因子，以及转录激活因子存在时前起始复合体的组装。除了图12-15中所示的通用转录因子之外，结合在基因附近位点的转录激活因子募集核小体修饰和重塑复合体、中介蛋白复合体，共同帮助形成前起始复合体。

除了在转录激活中的这一核心作用外，中介蛋白中单个亚基的缺失常常只导致一个小子集的基因失去表达，每个亚基对应不同子集的基因（它是由许多亚基构成的）。这一结果很可能反映了一个事实，即不同激活因子被认为与不同中介蛋白亚基相互作用而将聚合酶带到不同的基因上。此外，中介蛋白还通过调控TFIIH内的CTD激酶来帮助转录起始。

对核小体修饰因子和重塑因子的需求在不同的启动子上甚至同一启动子在不同的环境下都是不同的。无论何时何处，只要需要这些复合体也是由DNA结合的激活因子募集的。

我们将在第17章中讨论中介蛋白和修饰因子在激发转录中的作用。现在我们要考察中介蛋白的一些结构和功能上的特性。

中介蛋白包含很多亚基，其中有些亚基从酵母到人类是保守的

如图12-19所示，酵母和人的中介蛋白各自包括20个以上的亚基，其中7个在两种生物之间表现出显著的序列同源性（起初二者的亚基名字是不同的，反映出在确认它们的过程中所使用的实验方法的不同。但最近确立了一个协定，不同生物体中相当的亚基取一样的名字。图12-19采用了这种命名）。这些亚基的功能几乎都没有鉴定。只有一个（Srb4/Med17）是几乎所有Pol Ⅱ基因在细胞内转录所必需的。低分辨率的结构

比较提示两种中介蛋白具有相似的形状，而且二者都非常大，甚至比 RNA 聚合酶本身还要大。

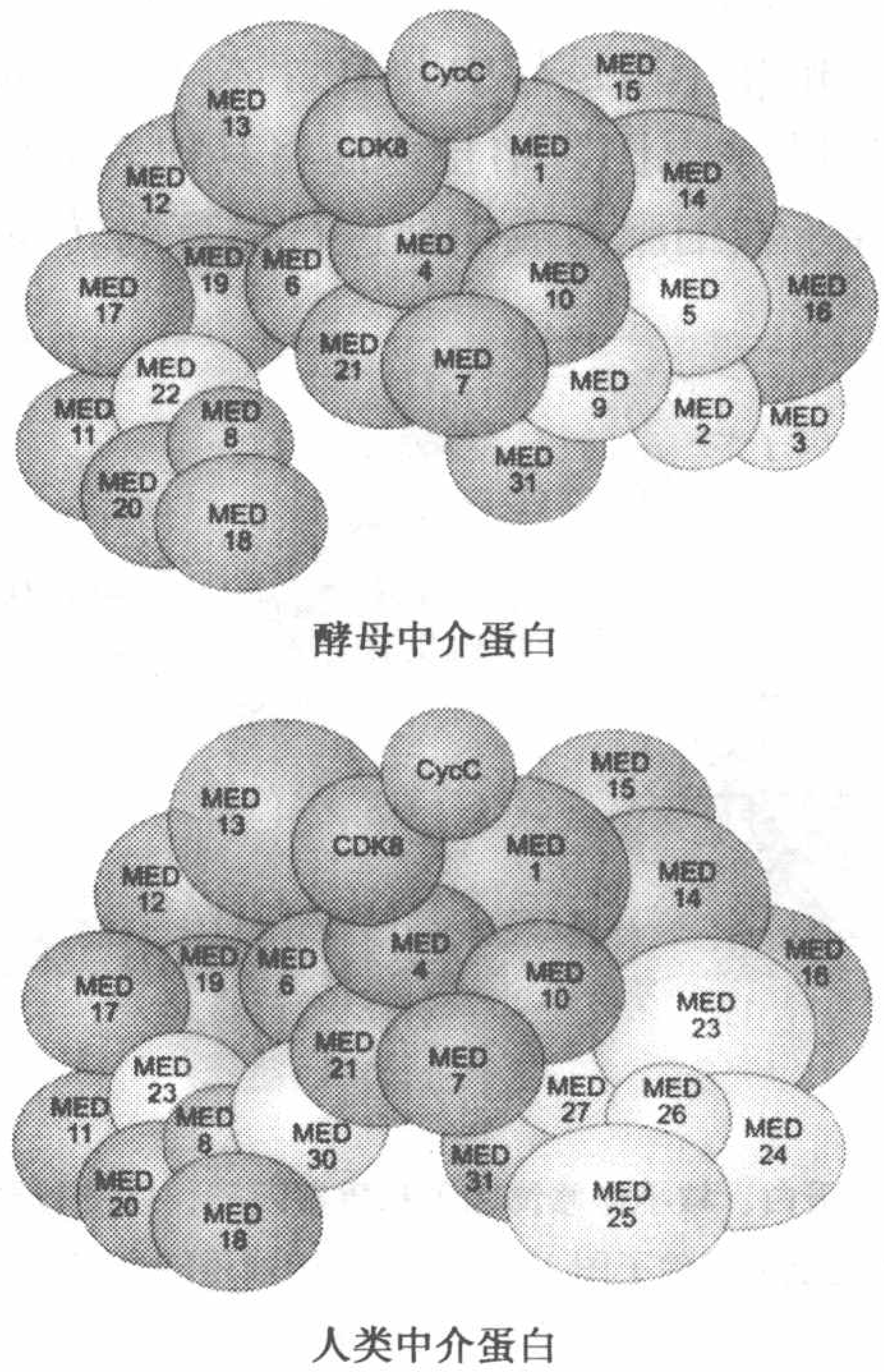


图 12-19　酵母和人类中介蛋白的比较。大多数亚基在酵母和人类中都有，不同的亚基以浅色阴影显示。（来源：Yeast Mediator：modified from Guglielmi B. et al. 2004. Nucleic Acids Res. 32：5379-5391，Fig. 8B；human Mediator：modified，with permission，from Malik S. and Roeder R. G. 2005. Trends biochem. Sci. 30：256—263，Fig. la. © Elsevier.）

酵母和人类的中介蛋白都是由一些模块组装而成的，每个模块是一组亚基（图 12-19）。在体外这些模块在某些条件下能够彼此分离。这一观测结果以及人类中介蛋白取决于分离手段可获得多种多样的组合（和大小）的事实，使人们认为存在各种形式的中介蛋白（尤其在多细胞动物中），每一种形式都包含一些中介蛋白亚基的子集。而且，还认为不同的形式参与调控不同子集的基因或者应答不同类别的调控因子（激活因子或抑制因子）。然而，同样可能的是所观察到的亚基组成的变化是人为造成的，仅仅反映了分离方法的不同。

一组新的因子激发 Pol Ⅱ 延伸和 RNA 校对

如前所述一旦聚合酶逃离了启动子并起始了转录，便转入了延伸阶段。这一转变包括 Pol Ⅱ 酶脱落其大部分起始因子，如通用转录因子和中介蛋白。另一组因子被募集在它们的位置上。其中的一些（如 TFIIS 和 hSPT5）是延伸因子（elongation factor），也就是激发延伸的因子，其他的是 RNA 加工所需要的。参与 RNA 加工的酶像一些之前

已经讨论过的起始因子一样，被募集到Pol Ⅱ大亚基的C端尾巴（CTD）上（图12-20）。然而，在这一情况中，那些因子更喜欢CTD被磷酸化的形式。因而，CTD的磷酸化导致了起始因子与延伸和RNA加工所需要的那些因子之间的交换。

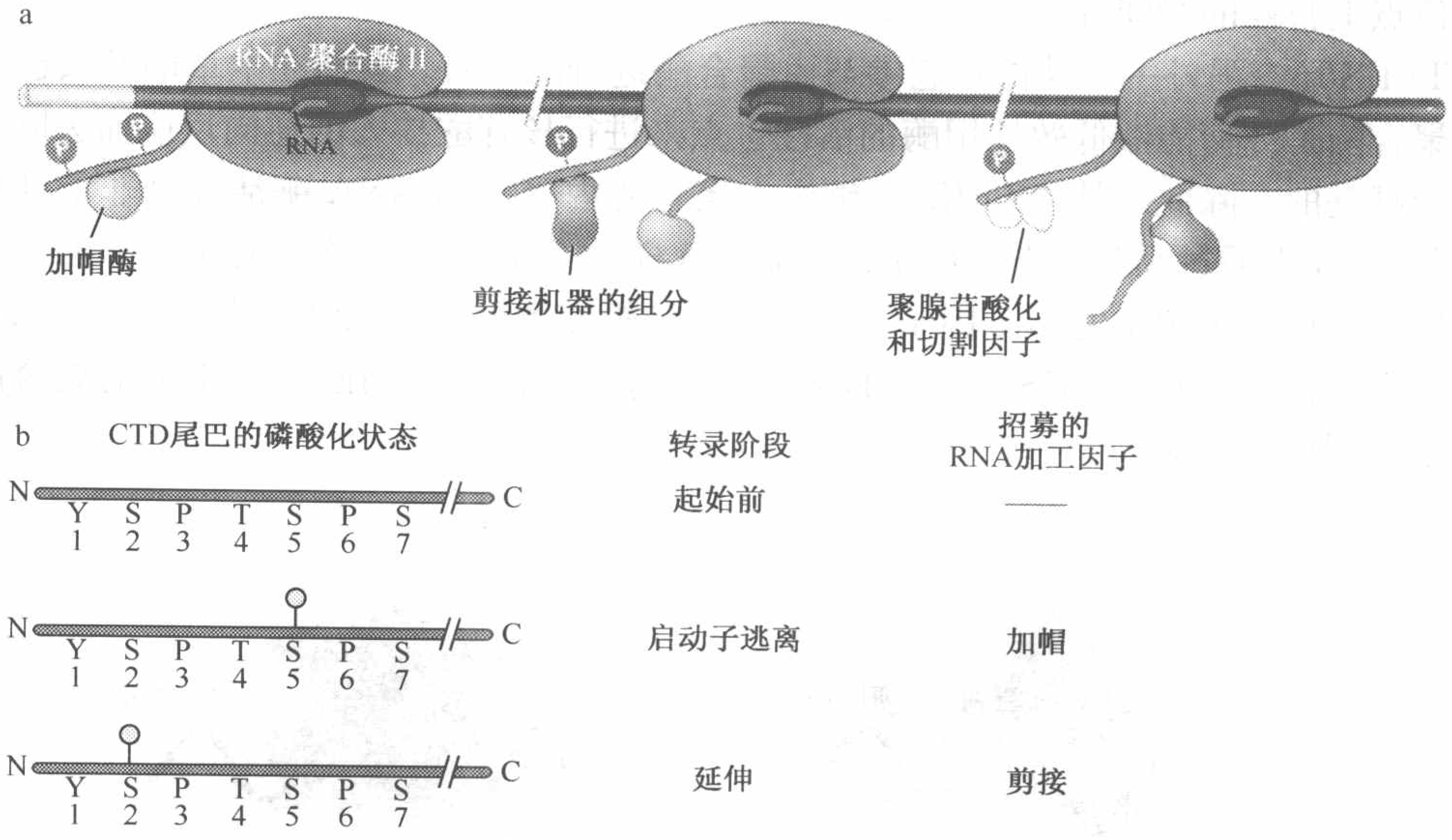


图 12-20 RNA 加工酶是由聚合酶的 CTD 尾巴募集的。 a. 参与 RNA 加工的各种因子由聚合酶的 CTD 尾巴募集。不同因子的募集取决于尾巴的磷酸化状态。这些因子随后在需要的时候转移到 RNA 上（见正文下节）。b. 第一行显示尾巴的示意图，以及一个拷贝的七肽重复序列。被磷酸化的丝氨酸残基的位置在第二、三行中标出。位置 5 处的丝氨酸残基的磷酸化见于启动子逃离后，并与加帽因子的募集相关。位置 2 处的丝氨酸残基的磷酸化见于延伸期间，并与剪接因子的募集相关。参与转录延伸和 RNA 加工的因子的募集是重叠的。因此，延伸因子 hSPT5 被募集到在 Ser5 磷酸化的尾巴上。

从酵母 Pol Ⅱ的晶体结构可以明显看出，聚合酶CTD的位置直接与新合成的RNA离开此酶的通道相邻。（它能够从酶体延伸出大约 800Å 它还非常长也就是说，约 7 倍于酶的其他部分的长度。）。这些特征使得它的尾巴可以与延伸和加工机器的几个组分结合并将它们运送到新生成的RNA处。

有多个蛋白质被确认可以激发由 Pol Ⅱ介导的延伸。其中之一是激酶 P-TEFb，它是由转录激活因子募集到聚合酶上的。一旦与 Pol Ⅱ结合，此蛋白质就将 CTD 重复序列中位置 2 处的丝氨酸残基磷酸化，此磷酸化事件与延伸相互关联。此外，P-TEFb磷酸化并因此激活另一个称为 hSPT5 的蛋白质，它本身是一个延伸因子。最后，还有另一个延伸因子，TAT-SF1，被 P-TEFb 募集。因而，P-TEFb 以三种不同的方式激发延伸。

另一类延伸因子是所谓的 ELL 家族。它们也结合延伸中的聚合酶并抑制聚合酶的瞬时停顿；否则的话，这种暂停会在 DNA 的许多位点发生。第一个人类 ELL 蛋白最初是在急性髓系白血病中作为一个经历置换的基因产物而被发现。

另一个不影响起始但激发延伸的因子是 TFIIS。类似 ELL 此因子促进延伸的总体

速度，这是通过在遭遇到趋向于减缓聚合酶进程的序列时限制其暂停时间而实现的。聚合酶并不是以一个固定的速度转录整个序列的，这是其特征之一。相反，它会周期性地暂停，有时在恢复转录前会暂停相当长的时间。在 TFIIS 存在时，聚合酶在任何一个特定位点上暂停的时间都会被缩短。

TFIIS 还有另外一个功能：它参与由聚合酶进行的校对。我们在本章的开头已经谈到了聚合酶是如何能够低效地用酶的活性位点来进行核苷酸掺入的逆向反应而去除错误加入的碱基的。除此之外，TFIIS 能够激发聚合酶内固有的 RNA 酶活性（非活性位点的部分），允许通过局部的有限的 RNA 降解这样一个不同的途径来去除错误加入的碱基。这一特征与细菌中的 Gre 因子激发的水解编辑相类似。图 12-21 显示了尽管结构上无关（序列也无关），TFIIS 和 Gre B 如何以同等的方式分别与酵母和细菌的聚合酶结合来激发同样的反应。

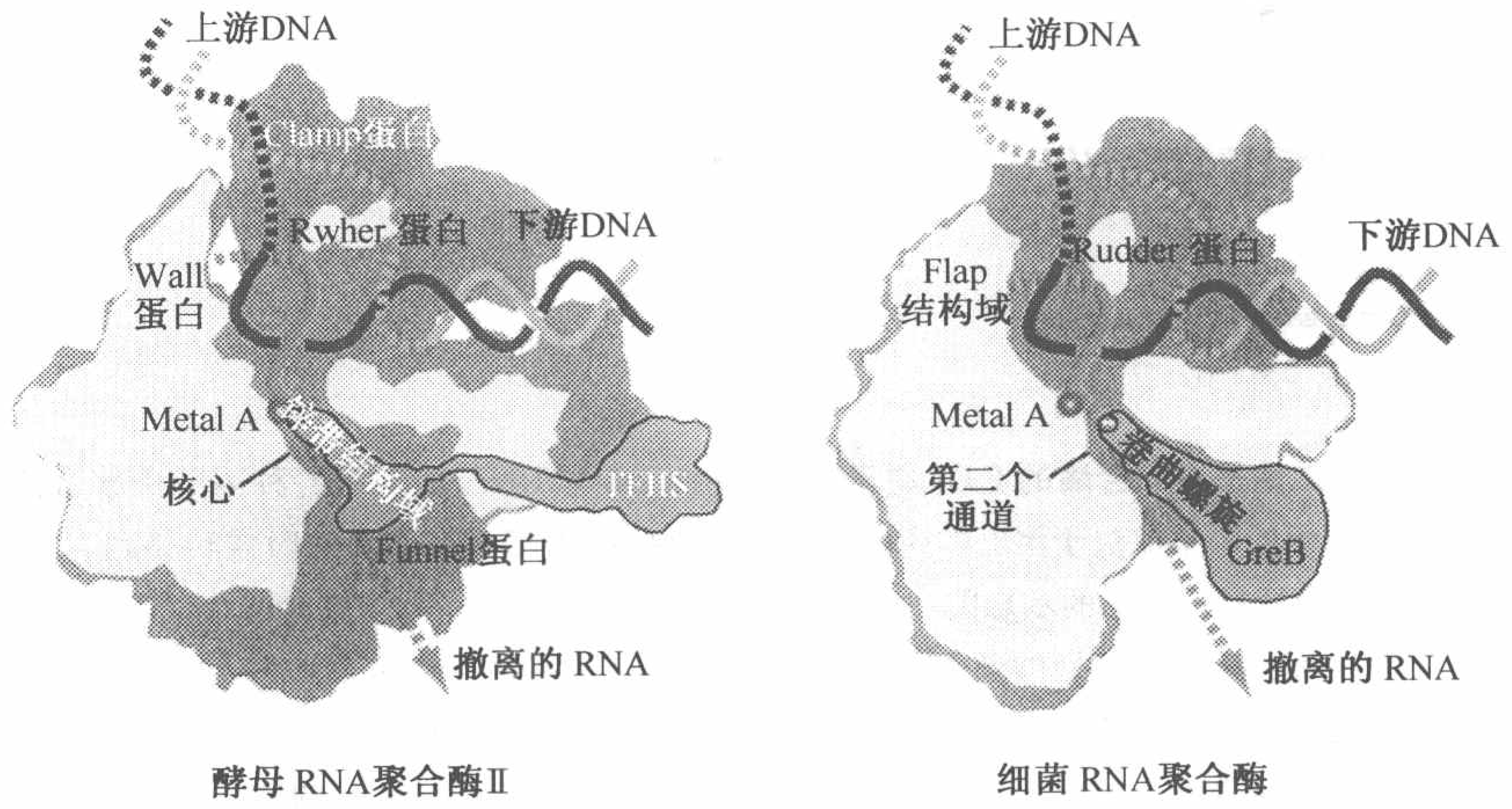


图 12-21 TFIIS 和 GreB 行为相似。本图以切掉一部分复合体的形式显示暂停的 RNA 聚合酶Ⅱ与 TFIIS 复合体（左图）及细菌 RNA 聚合酶与 GreB（右图）的主要特征。TFIIS（橙色）插入 RNA 聚合酶Ⅱ的核心，而 GreB（也是橙色）插入细菌 RNA 聚合酶的通道。在每种情况下，主要的催化镁离子指定为 Metal A（粉色），两个保守的酸性残基用绿色环表示。因此我们看见虽然这两个蛋白是这么不同，它们的行为基本上一样。虚线箭头表示撤离的 RNA 的假定位置（也参见图 12-10）。（来源：Conaway R. C. et al. 2003. Cell 114：272-274，Fig. 1. © Elsevier. 经许可复制）

延伸 RNA 聚合酶在行进过程中必须应对组蛋白

与转录起始一样，延伸也在组蛋白的存在下进行，因为 DNA 模板整合在核小体内。RNA 聚合酶如何通过这些潜在的障碍而转录？

比较裸露 DNA 和整合在染色质中的 DNA 的体外转录实验揭示了染色质极大地阻碍转录。这个实验设置提供了在染色质存在的情况下确定促转录因子的分析方法。如此，在人类细胞提物中确定了一个称为 FACT（facilitates chromatin transcription）的因子。顾名思义，这个因子使在染色质模板进行的转录更有效。FACT 是两个非常保守

的蛋白的异型双聚体；Spt16 和 SSRP1。前者的酵母同源基因在遗传学研究中已经被确定与染色质调制有关。FACT 在延伸中的作用是通过这个复合体与已知的延伸因子 TFIIS 之间的遗传相互作用而建立的。

FACT 如何起作用？第 7 章提到核小体是八聚体，由 H2A、H2B、H3、H4 组蛋白和 DNA 组成（图 7-20）。这些组蛋白排成两个模块：H2A/H2B 二聚体和 H3/H4 四聚体。Spt16 结合前者：SSRP1 结合后者。令人惊奇的是，FACT 可以通过挪开一个 H2A/H2B 二聚体而拆除组蛋白，也可以通过放回 H2A/H2B 二聚体而重新组装组蛋白。

因此，我们描绘出在延伸过程中 FACT 如何起作用的画面（图 12-22）。在转录中的 RNA 聚合酶的前方，FACT 挪开一个 H2A/H2B 二聚体。这使得聚合酶通过那个核小体（在体外，已经显示了从模板上挪开一个 H2A/H2B 二聚体使转录能够进行）。FACT 还有组蛋白伴护活性；这使得它能够在行进的聚合酶后将 H2A/H2B 二聚体立即放回到组蛋白六聚体中。如此，FACT 使得聚合酶延伸，同时保持了转录中的染色质的完整性。

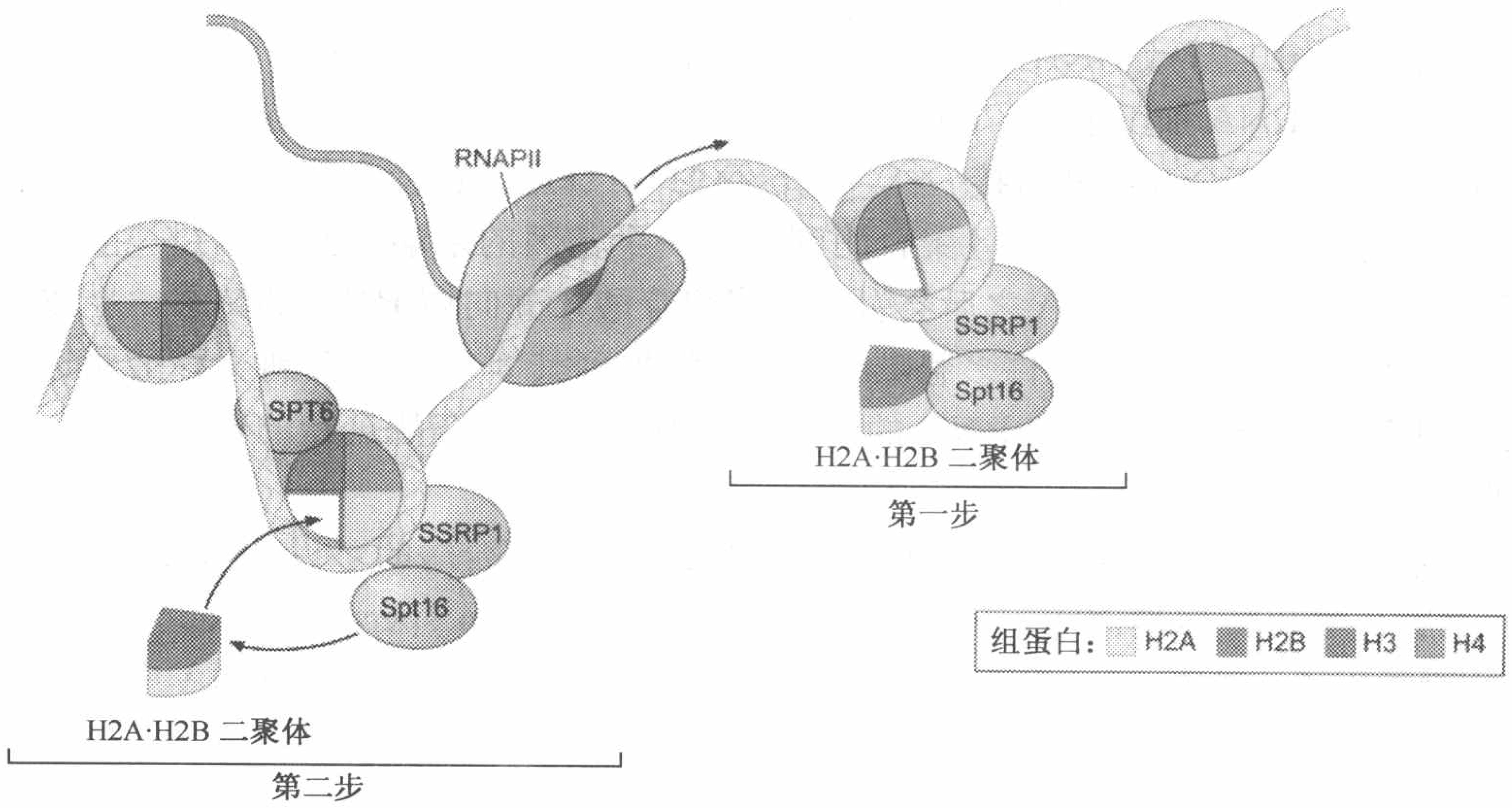


图 12-22 FACT 协助的通过核小体的延伸模型。如文中所述，FACT. 显示为 Spt16 和 SSRP1 的异型二聚体，能够在转录的 RNA 聚合酶前方拆除核小体（步聚一）并在其后重装（步骤二）。具体地说，它挪开 H2A/H2B 二聚体。STP6 结合组蛋白 H3，相信能帮助核小体重装。（来源：Reinberg D. and Sims R. 2006. J. Biol. Chem. 281：23297-23301，Fig. 2b. © American Society for Biochemistry and Molecular Bilogy. 经许可重绘）

延伸聚合酶与各种 RNA 加工所需要的一组新的蛋白质因子相互作用

一旦被转录，真核细胞的 RNA 必须以各种方式进行加工，然后再从细胞核输出到可翻译的地方。这些加工事件包括下面几个：RNA5′端加帽（capping）、剪接（splicing）和 RNA3′端的多聚腺苷酸化（polyadenylation）。其中最复杂的是剪接——非编码的内含子从 RNA 中去除而产生成熟的 mRNA 的过程。这一过程和其他过程（如 RNA 编辑）的机制和调控是第 13 章的主题。我们在这里讨论另外两个过程。

引人注目的是，参与延伸的蛋白质和 RNA 加工所需要的蛋白质存在交叠。例如，上面提到过的一个延伸因子（hSPT5）也帮助募集 5′加帽酶到聚合酶的 CTD 尾巴（位置 S 上的丝氨酸磷酸化（图 12-20b））。hSPT5 激活 5′加帽酶的活性。又如，延伸因子 TAT-SF1募集剪接机器的组分到带着 Ser2 磷酸化尾巴的聚合酶上（图 12-20b）。因而转录的延伸、终止和 RNA 加工似乎是相互联系的，人们推测这是为了保证它们的正确配合。

第一个 RNA 加工事件是加帽，涉及在 RNA 的 5′端添加一个修饰过的鸟嘌呤碱基。具体来讲，这是一个甲基化的鸟嘌呤，而且以一种不常见的包括 3 个磷酸盐的 5′-5′连接方式加入到 RNA 转录物中的（图 12-23 的底部）。

5′帽子是由 3 个酶反应步骤生成的，细节见图 12-23 及其说明。在第一个步骤中，一个磷酸盐基团从转录物的 5′端去除。然后在第二个步骤中，GMP 部分加上。在最后一步中，给此核苷酸添加上一个甲基。RNA 从聚合酶的 RNA 出口通道一出现就给加帽了。这个发生在转录周期刚刚进入了起始和延伸阶段间的转变。加帽之后，尾部序列重复中的 Ser5 发生去磷酸化可能导致加帽机器的脱离，而进一步的磷酸化（这次是尾部重复中的 Ser2）导致 RNA 剪接所需要的机器的募集（图 12-20b）。

最后一个 RNA 加工事件，mRNA 3′端的多聚腺苷酸化，是与转录的终止密切相关的（图 12-24）。正如加帽和剪接一样，聚合酶的 CTD 尾巴也参与募集某些多聚腺苷酸化所必需的酶（图 12-20）。一旦聚合酶到达了一个基因的末端，就会遇到一段特殊的序列，此序列在转录到 RNA 之后会引发多聚腺苷酸化酶向该 RNA 的转移，导致 4 个事件的发生：mRNA 切割、许多腺嘌呤残基被添加到其 3′端、由 5′到 3′核酸酶进行的对还与 RNA 聚合酶结合着的 RNA 的降解、以及随后由聚合酶进行的转录终止。下面详述这一系列事件。

在聚合酶接近基因的末端时，两个蛋白质复合体由聚合酶的 CTD 运载：CPSF（切割和聚腺苷酸化特异性因子，cleavage and polyadenylation specificity factor）和 CstF（切割刺激因子，cleavage stimulation factor）。有些序列一旦转录到 RNA 中，就会引发这些因子向 RNA 的转移，这些序列称为 poly-A 信号，它们的作用方式在图 12-24 中显示。一旦 CPSF 和 CstF 结合到 RNA 上，其他蛋白质也就被募集，先引起 RNA 的切割，然后是多聚腺苷酸化。

多聚腺苷酸化是由 poly-A 聚合酶介导的，此酶在由切割所形成的 RNA 的 3′端添加大约 200 个腺嘌呤。它以 ATP 作为前体并以与 RNA 聚合酶相同的化学机制添加核苷酸，但是它这样做并不需要模板，因而，长腺苷酸尾巴见于 RNA 中，而非 DNA 中。目前还不清楚决定 poly-A 尾巴的长度的因素是什么，但是该过程涉及其他与 poly-A 序列特异结合的蛋白质，然后成熟的 mRNA 从细胞核中运出（在第 13 章中讨论）。值得注意的是，长腺苷酸尾巴是由 Pol Ⅱ产生的转录物所特有的，此特征使得可以在实验中用亲和层析法分离蛋白质编码的 mRNA。

如此，我们了解了一旦基因被转录，成熟的 mRNA 是如何从聚合酶上释放的。但是是什么终止聚合酶的转录呢？实际上，聚合酶并不在 RNA 被切割和多聚腺苷酸化时就立刻终止。相反，它继续沿着模板移动，产生第二个 RNA 分子。聚合酶在终止前继

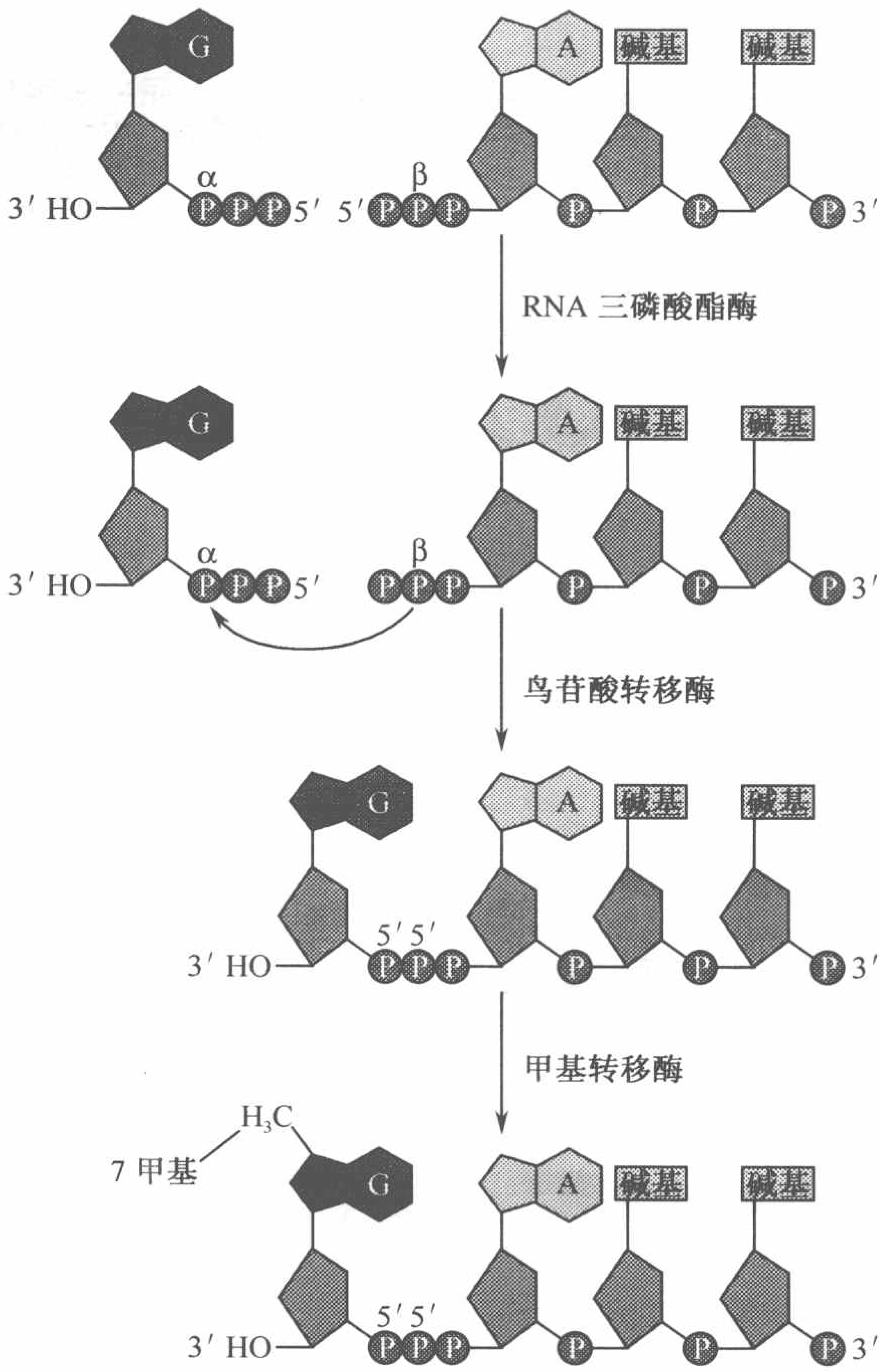


图 12-23 RNA 的 5′帽子的结构和形成。在第一个步骤中，RNA 5′端的 γ-磷酸盐由 RNA 三磷酸酯酶（triphosphatase）去除（一个转录物的起始核苷酸最初保留着其 α-、β-和 γ-磷酸盐）。在下一个步骤中，鸟苷酸转移酶（guanylyl transferase）把 GMP 部分加到末端的 β-磷酸盐上。这是一个两步过程：首先，酶— GMP 复合体通过 GTP 释放 β-和 γ-磷酸盐而产生。然而酶的 GMP 转移到 RNA5′端的 β-磷酸盐上。一旦形成了这一连接，新加上的鸟嘌呤和 mRNA 起初的 5′端的嘌呤就由甲基转移酶加上甲基基团而得到进一步的修饰。所形成的 5′帽子结构随后将核糖体募集到 mRNA 上以准备开始转录（见第 14 章）。

续转录的 4 个核苷酸才从模板上分离现在描述现行的终止发生的模型。

转录终止与高活性的 RNase 实施的对 RNA 的毁坏紧密关联

多聚腺苷酸化与转录终止密切相关，虽然到底是如何关联的还不十分清楚。最近，当第二个 RNA 从聚合酶出现就降解它的酶已经被确认，这个酶本身可能就引发终止。这被称为终止的“鱼雷模型（torpedo model)”（图 12-25a）。

第二个 RNA 的自由端没有被加帽，因此可以与真正的转录物加以区别。这个新 RNA 被一种在酵母中称为 Rat1（人类中是 Xrn2）的 RNase 所识别，该酶被另一个与

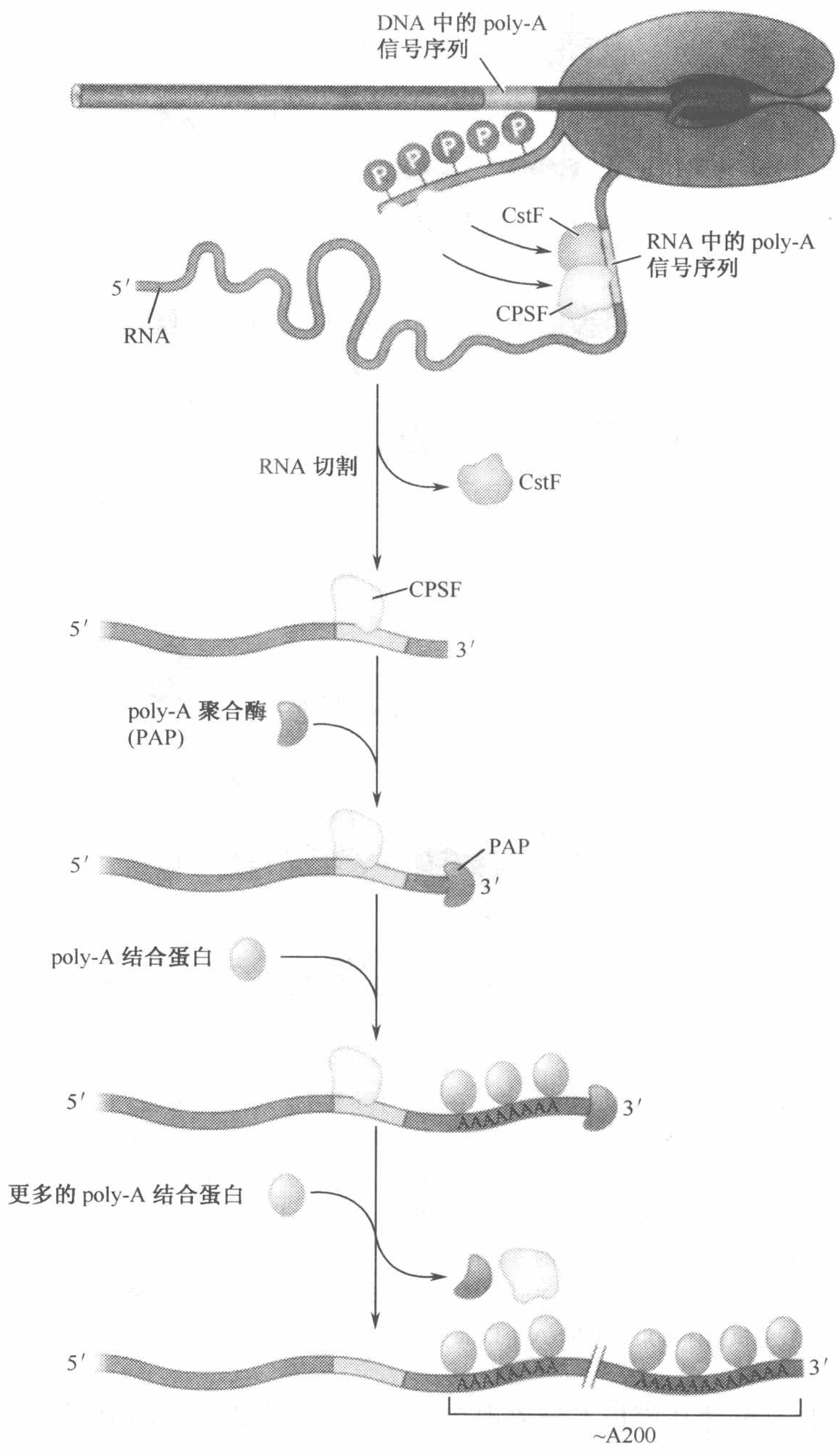


图 12-24 多聚腺苷酸化和终止。此过程中的各个步骤在文中有描述。

RNA 聚合酶的 CTD 结合的蛋白（Rtt103）运送到 RNA 的末端。Rat1 酶活性很高，从 5′→3′方向迅速地降解 RNA，直到追上还在转录的、RNA 不断从中涌出的聚合酶。转录终止可能不需要 Rat1 和聚合酶之间有任何非常特异的相互作用，可能实际上是以与本章前面部分描述过的细菌中 Rho-依赖型终止类似的方似引发的，即高活性的

RNase 把聚合酶往前推并/或者将余下的新生 RNA 转录物从聚合酶中拉出来。

尽管转录终止的鱼雷模型在目前占优势，但还是有另外一个模型称为“变构模型(allosteric model)”（图 12-25b）。根据这个模型，终止取决于延伸聚合酶的构象改变，构象改变降低了酶的活性，导致其后不久就自发终止。这个构象改变可能与多聚腺苷酸化相关联，可能的引发原因有聚合酶 CTD 尾巴的 3′加工酶转移到 RNA、随后结合到其他诱发构象改变的因子的 CTD 尾巴上等。

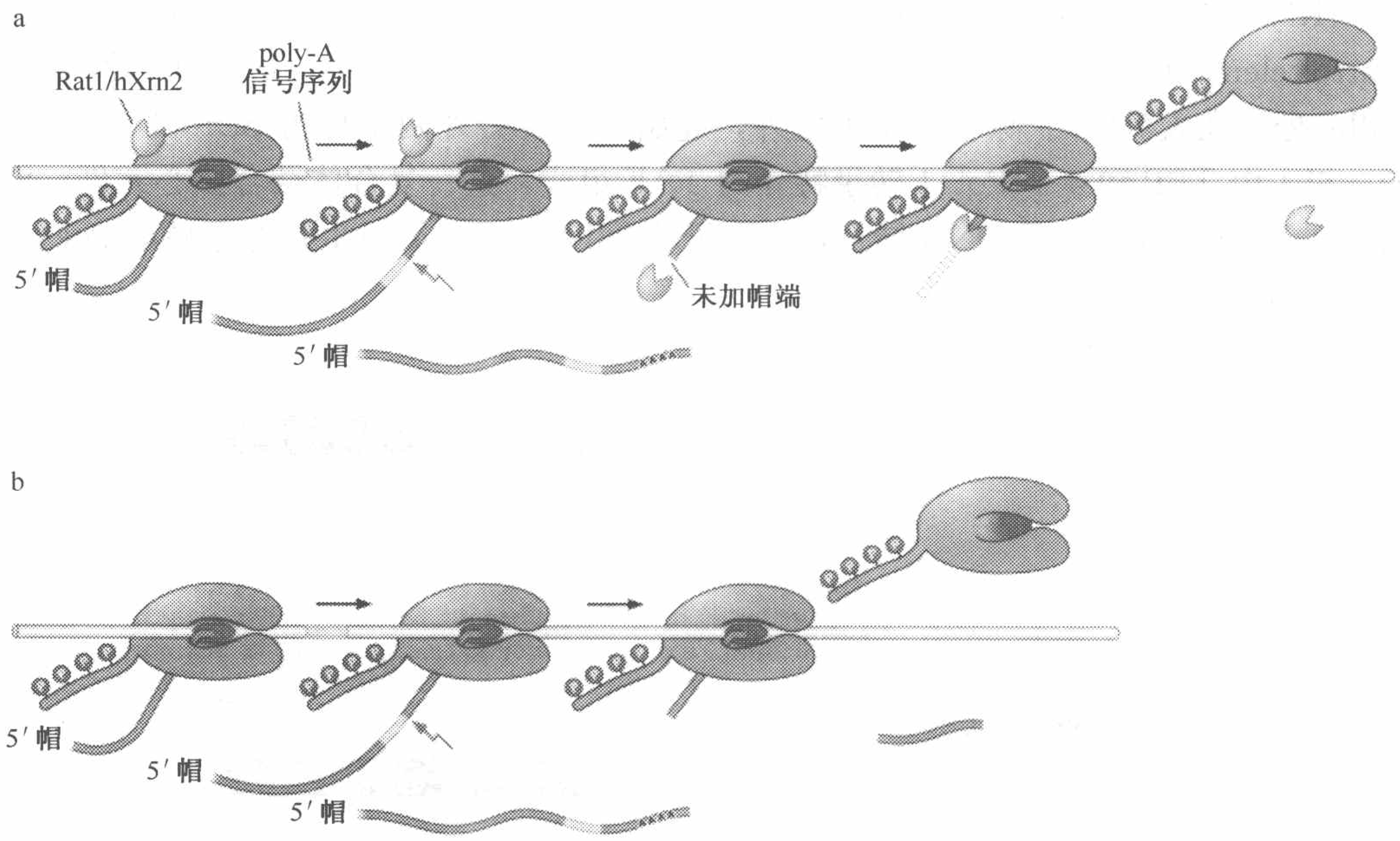


图 12-25 转录终止的模型：鱼雷（torpedo）和变构（allosteric）。如正文中所述，对于真核 RNAPol Ⅱ转录一个基因后是如何终止的存在两个模型。图中，在 DNS 上的 poly-A 位点就在紧邻基因的下游，用亮绿段表示。在转录物中也用亮绿表示。绿色虚线表示降解的转录物。a. 在鱼雷模型中，poly-A 位点下游转录的 RNA 被 5′→3′ RNase（鱼雷）攻击，该 RNase 由聚合酶本身带到这个转录物上。当这个核酸外切酶追上聚合酶时，它引发了从 DNA 模板上解离和转录的终止。b. 在变构模型中，聚合酶在转录基因时具有高活性，然后，一旦过了 poly-A 位点，活性降低。这个改变可能由于修饰和构象改变。即使在变构模型，第二个 RNA 要被 RNase 降解，但不是终止的原因。在此情况下，RNA 降解没有在图中显示，以强调这两个模型不同的终止机制。（来源：Adapted，with permission，from Luo W. and Bentley D. 2004. Cell 119：911-914，Fig. 1. © Elsevier.）

由 RNA 聚合酶Ⅰ和Ⅲ催化的转录

RNA 聚合酶Ⅰ和Ⅲ识别不同的启动子，使用不同组别的转录因子，但还是需要 TBP

我们已经提到，真核细胞中除 Pol Ⅱ之外还有另外两个聚合酶，Pol Ⅰ和 Pol Ⅲ。这些酶是与 Pol Ⅱ相关的，甚至共用几个亚基（表 12-2），但是它们从不同的启动子起始转录并转

录不同的基因。这些基因编码专门的 RNA，而不是蛋白质，这一点我们已经在本章前面的部分中讨论过。这些酶也各自使用自己独特的一组通用转录因子。然而，TBP 是通用的，它不仅参与 Pol Ⅱ起始的转录，也参与 Pol Ⅰ和 Pol Ⅲ的转录。

Pol Ⅰ只用于一个基因的表达，那就是合成 rRNA 前体的基因。该基因在每个细胞中都有很多拷贝，事实上它的表达水平远远高于其他任何基因，这也许解释了为什么它有自己专用的聚合酶。

rRNA 基因的启动子包括两个部分：核心元件和 UCE（上游控制元件，upstream control element），如图 12-26 所示。前者位于转录起始位点周围，后者在上游 100～150bp 处（在人类中）。除了 Pol Ⅰ之外，起始还需要另外两个称为 SL1 和 UBF 的因子。SL1 由 TBP 和 3 个 Pol Ⅰ转录特异的 TAF 组成，这一复合体结合核心元件。SL1 只有在 UBF 存在时才与 DNA 结合，该因子结合 UCE，引入 SL1 并通过募集 Pol Ⅰ而从核心启动子激发转录。

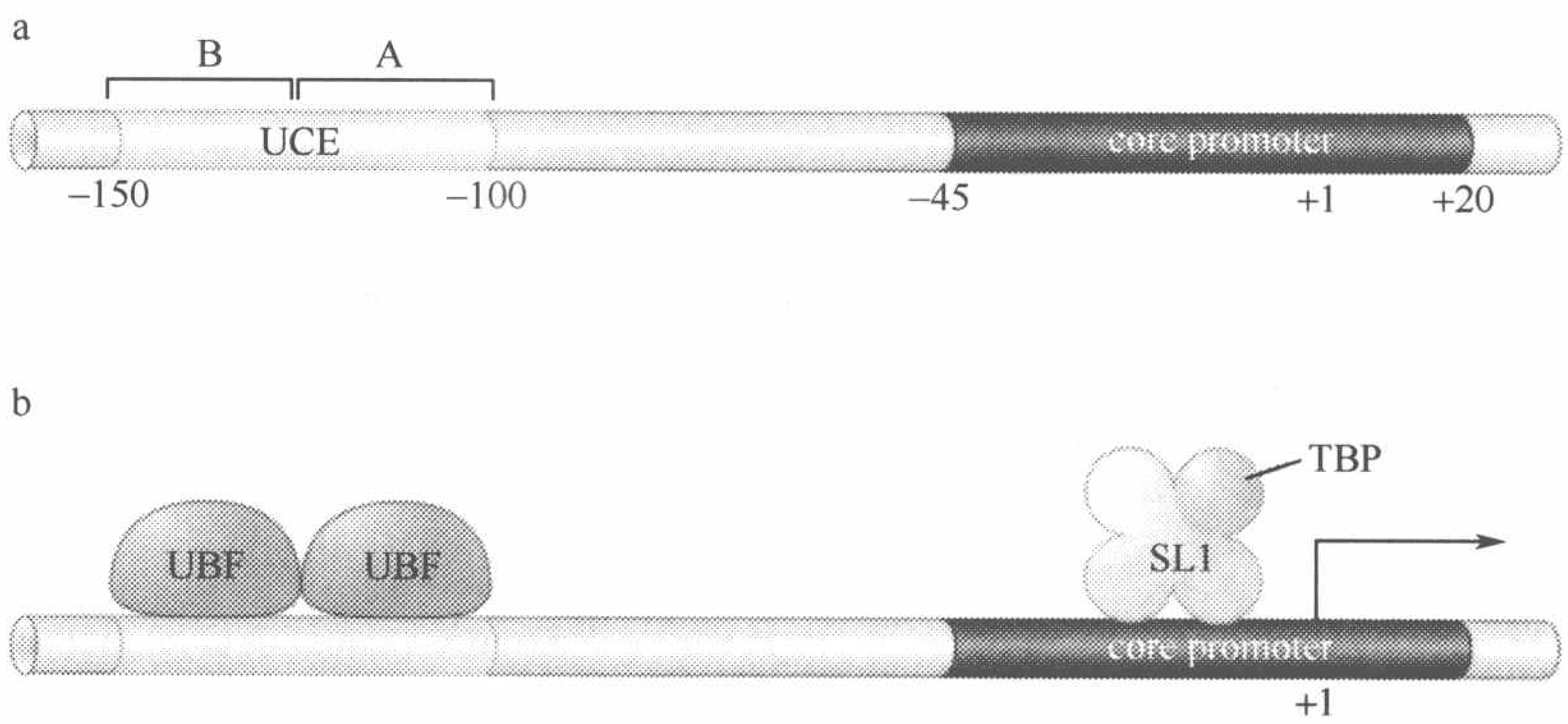


图 12-26 Pol Ⅰ启动子区域。 a. Pol Ⅰ启动子的结构；b. Pol Ⅰ转录因子。这里显示的是人类的情况。酵母中参与帮助 Pol Ⅰ转录的一组蛋白质是明显不同的。

PolⅢ启动子位于转录起始位点的下游

Pol Ⅲ启动子有各种形式，而且绝大部分具有位于转录起始位点下游的不寻常特征。一些 Pol Ⅲ启动子（如 tRNA 基因的启动子）包含两个区域，称为盒子 A 和盒子 B，二者被一个短的元件分隔开（图 12-27）；另一些包含盒子 A 和盒子 C（如 5S rRNA 基因）；还有一些包含与 Pol Ⅱ启动子相似的 TATA 元件。

与 Pol Ⅱ和 Pol Ⅰ很相似，Pol Ⅲ进行的转录除需要聚合酶之外还需要转录因子。在这一情况中，这些因子称为 TFIIIB 和 TFIIIC（对于 tRNA 基因），5S rRNA 基因的转录还需要加上 TFIIIA。

图 12-27 显示了 tRNA 的启动子。这里，TFIIIC 复合体结合到启动子区域。这个复合体将 TFIIIB 募集到紧邻起始位点的上游 DNA 上，在那里它转而将 Pol Ⅲ募集到转录的起始位点上。然后此酶起始转录，推测当它转录时将 TFIIIC 从 DNA 模板上替换掉。与另两类聚合酶一样，Pol Ⅲ也使用 TBP。在这一情况中，那个普遍存在的因子见于 TFIIIB 复合体中。

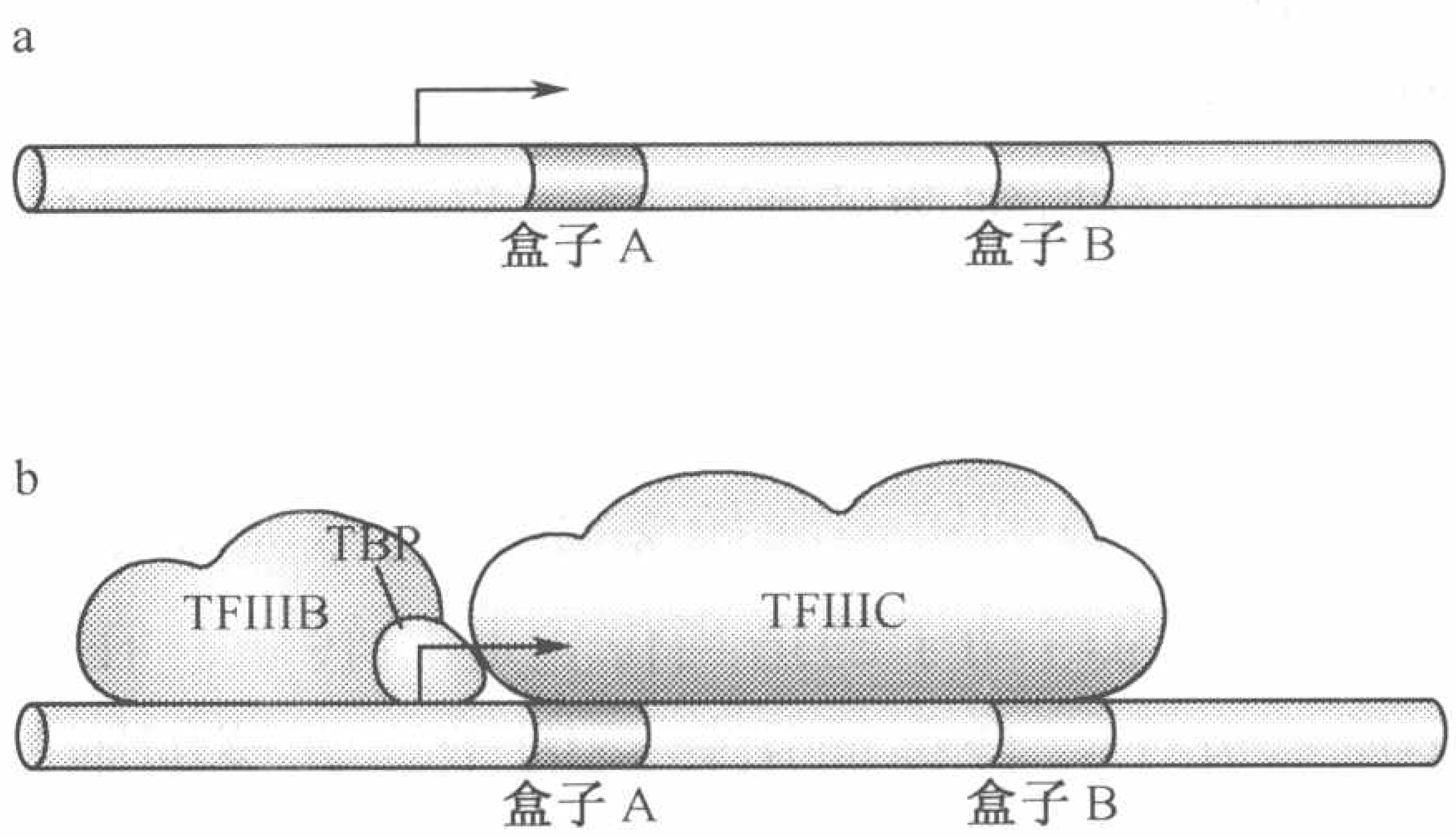


图 12-27 Pol Ⅲ核心启动子。这里显示的是一个酵母 tRNA 基因的启动子。导致转录起始的事件的次序在正文中有描述。

小结

基因表达是 DNA 双螺旋中的信息转移到 RNA 和蛋白质中的过程，蛋白质的活性则赋予了一个细胞其形态学和功能。转录是基因表达的第一步，包括将 DNA 复制成 RNA。由 RNA 聚合酶催化的这一过程在很多方面都与第 8 章中讨论的 DNA 复制的过程相似。在这两种情况中，都从一条 DNA 模板上合成一条新的核苷酸链，并且 DNA 和 RNA 的合成也都是以 5′→3′方向进行的（也就是说，酶是把每个相继的核苷酸添加到正在延长的链的 3′端）。但是这两个过程之间存在几个关键的区别，一些是机制上的，另一些则反映它们所起的不同作用。

例如，在 DNA 复制中，整个基因组被复制一次，而且每次细胞分裂只复制一次。在转录中，基因组中只有一些区域得到转录，而且这些选中的区域在不同的细胞中或同一细胞的不同时期中是不同的。不同的区域可转录的程度也不同，即在单个细胞里，一个特定的区域可生成从一个到几千个转录物。

转录和复制在机制上的区别包括以下方面：用于生成一条新 DNA 链的核苷酸是脱氧核糖核苷酸，而在转录中则是核糖核苷酸。还有，DNA 聚合酶只能延伸已经存在的多聚核苷酸链，因而需要一段引物，而 RNA 聚合酶能够从头起始 RNA 的合成。

RNA 聚合酶从细菌到人是高度保守的。真核细胞有 3 个不同的聚合酶，细菌只有一个。真核细胞的 3 个酶称为 RNA Pol Ⅰ、Ⅱ和Ⅲ。在本章中我们主要集中在Pol Ⅱ上，因为它是转录细胞中绝大部分基因和全部蛋白质编码基因的酶。

大肠杆菌的基本酶称为核心酶，含有各一个拷贝的 β、β′和 ω 亚基和两个拷贝的 α 亚基。所有这些亚基在真核细胞的酶中都有同源物。细菌和酵母 Pol Ⅱ的结构也很相似，二者都组装成一个蟹爪的形状，在细菌的酶中钳子是由最大的亚基 β 和 β′组成的。活性位点在钳子的基部，进出活性位点的途径是由 5 个通道提供的：一个允许双链 DNA 从酶前部的钳子之间进入；另外两个允许两条单链——模板和非模板链通过活性

位点之后离开酶；另一个通道为 NTP 进入活性位点提供途径；在聚合位点后面一小段距离从 DNA 模板上剥离下来的 RNA 产物，则从第 5 个通道离开酶。

Pol Ⅱ在一个很重要的方面与细菌的酶不同。前者在大亚基的 C 端末尾有一个所谓的“尾巴”，而细菌的酶是没有的。这个尾巴是由许多七肽序列重复组成的。

一个周期的转录要经过 3 个阶段，称为起始、延伸和终止。虽然 RNA 聚合酶能够独立地合成 RNA，但精确而有效的转录起始需要其他称为起始因子蛋白质的参与。这些因子确保酶只从 DNA 上称为启动子的适当位置起始转录。在细菌中只有一个起始因子 σ，而在真核细胞中则有几个，合称通用转录因子。真核细胞的 DNA 是包裹在核小体内的，在细胞内的有效起始通常需要额外的蛋白质参与，包括中介蛋白复合体和核小体修饰酶。转录激活蛋白也是需要的（第 17 章）。

在起始过程中，RNA 聚合酶（和起始因子一起）在一个闭合式复合体中与启动子结合。在该状态下 DNA 仍保持着双链的形式，然后这一闭合式复合体经历异构化作用而变为开放式复合体。在该形式中，转录起始位点周围的 DNA 解开，从而破坏碱基对并形成一个单链 DNA 泡。这一转变使得聚合酶等可以接近决定新 RNA 链中碱基顺序的模板链。起始的这一阶段之后是启动子逃离：一旦酶合成了一系列短 RNA（称为流产性起始），它会设法生成一段超过 10 个碱基的转录物。此时酶离开启动子而进入了延伸阶段。在这一阶段中，聚合酶沿着基因移动，同时执行几项功能：它打开 DNA 的下游序列并在活性位点的上游（后方）将其重新合上；它将核糖核苷酸添加到正在延长的转录物的 3′端；它在聚合位点后方大约 8bp 或 9bp 的位置将新形成的 RNA 链从模板上剥离；它还校对转录物，查找（并替换）错误插入的核苷酸。

细菌和真核细胞中的转录按照这些相同的步骤进行。然而，两者也存在一些区别。例如，在细菌中，向开放式复合体转变的异构化作用是自发的，不需要 ATP 水解；在真核细胞中这一步骤则是需要 ATP 水解的。更为惊人的是，在真核细胞中启动子逃离是由 CTD 尾巴的磷酸化状态调控的。这样。在前起始复合体中与启动子结合的 Pol Ⅱ有一个未磷酸化的 CTD。CTD 被一个或多个激酶磷酸化，其中一个是通用转录因子之一（TFIIH）的一部分。

一旦被磷酸化，Pol Ⅱ的 CTD 尾巴就将自己从与启动子上的其他蛋白质的结合状态中释放出来，使聚合酶进入延伸阶段。然后 CTD 与参与转录延伸和 RNA 加工的因子相结合。因而，当聚合酶从启动子移走并开始转录基因时，存在一个起始因子与延伸和加工因子的互换。延伸因子和参与加工的因子之间也存在相互作用，确保这些事件的正确协调。细菌和真核之间的另一个不同点是后者必须在延伸阶段对付核小体。这要求另一个复合体，可以在前方拆除核小体，在后方重装核小体，使聚合酶向前。

在细菌和真核细胞中转录终止也是不同的。在细菌中有两种终止子——固有（Rho-非依赖型）和 Rho-依赖型。固有终止子包括两个转录到 RNA 后才起作用的序列元件。一个元件是一段在 RNA 内形成一个茎-环结构的反向重复序列，破坏延伸聚合酶。与一串 U 核苷酸（仅与模板链微弱结合）共同作用，引起转录物的释放。Rho-依赖型终止子需要 ATP 酶 Rho，一个跳上正在延伸的转录物上并沿着转录移动直到遇上聚合酶引发终止的蛋白质。但在这些生物体中，转录终止据信涉及另一个蛋白，

RNase，沿着一条新生的转录物行进直到撞到聚合酶，引发终止。在真核细胞中，终止是与 3′聚腺苷酸化的 RNA 加工事件密切相关的。

在本章中我们讨论了 RNA 转录物 5′端的加帽、3′端的聚腺苷酸化，以及后者与转录终止之间的联系。剪接将在下一章中进行描述。

（胡少晖 吴琳 译 吴琳 校）

参考文献

书

Cold Spring Harbor Symposia on Quantitative Biology. 1998. Volume 63: Mechanisms of transcription. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Ptashne M. and Gann A. 2002. *Genes and signals.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

White R.J. 2001. *Gene transcription: Mechanisms and control.* Blackwell Science, Malden, Connecticut.

RNA聚合酶

Cramer P. 2004. RNA polymerase II structure: From core to functional complexes. *Curr. Opin. Genet. Dev.* **14:** 218–226.

Darst S. 2004. New inhibitors targeting bacterial RNA polymerase. *Trends Biochem. Sci.* **29:** 159–162.

Ebright R.H. 2000. RNA polymerase: Structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *J. Mol. Biol.* **304:** 687–698.

Murakami K.S. and Darst S.A. 2003. Bacterial RNA polymerases: The whole story. *Curr. Opin. Struct. Biol.* **13:** 31–39.

Paget M.S. and Helmann J.D. 2003. The σ70 family of sigma factors. *Genome Biol.* **4:** 203.

启动子

Butler J.E. and Kadonaga J.T. 2002. The RNA polymerase II core promoter: A key component in the regulation of gene expression. *Genes Dev.* **16:** 2583–2592.

Juven-Gershon T., Hsu J.-Y., and Kadonaga J.T. 2006. Perspectives on the RNA polymerase II core promoter. *Biochem. Soc. Trans.* **34:** 1047–1050.

转录起始

Boeger H., Bushnell D.A., Davis R., Griesenbeck J., Lorch Y., Strattan J.S., Westover K.D., and Kornberg R.D. 2005. Structural basis of eukaryotic gene transcription. *FEBS Lett.* **579:** 899–903.

Conaway R.C., Sato S., Tomomori-Sato C., Yao T., and Conaway J.W. 2005. The mammalian Mediator complex and its role in transcriptional regulation. *Trends Biochem. Sci.* **30:** 250–255.

Malik S. and Roeder R.G. 2005. Dynamic regulation of pol II transcription by the mammalian Mediator complex. *Trends Biochem. Sci.* **30:** 256–263.

Myers L.C. and Kornberg R.D. 2000. Mediator of transcriptional regulation. *Annu. Rev. Biochem.* **69:** 729–749.

Roberts J.W. 2006. RNA polymerase, a scrunching machine. *Science* **314:** 1097–1098.

Woychik N.A. and Hampsey M. 2002. The RNA polymerase II machinery: Structure illuminates function. *Cell* **108:** 453–463.

Young B.A., Gruber T.M., and Gross C.A. 2002. Views of transcription initiation. *Cell* **109:** 417–420.

延伸及RNA加工

Armache K.-J., Kettenberger H., and Cramer P. 2005. The dynamic machinery of mRNA elongation. *Curr. Opin. Struct. Biol.* **15:** 197–203.

Bentley D. 2005. Rules of engagement: Co-transcriptional recruitment of pre-mRNA processing factors. *Curr. Opin. Cell. Biol.* **17:** 251–256

Maniatis T. and Reed R. 2002. An extensive network of coupling among gene expression machines. *Nature* **416:** 499–506.

Reinberg D. and Smith R.J., III 2006. de FACTo nucleosome dynamics. *J. Biol. Chem.* **281:** 23297–23301.

Saunders A., Core J.C., and Lis J.T. 2006. Breaking barriers to transcription elongation. *Nat. Rev. Mol. Cell. Biol.* **7:** 557.

终止

Luo W. and Bartley D. 2004. A ribonucleolytic rat torpedoes RNA polymerase II. *Cell* **119:** 911–914.

Richardson J.P. 2003. Loading Rho to terminate transcription. *Cell* **114:** 157–159.

———. 2006. How Rho exerts its muscle as RNA. *Mol. Cell* **23:** 711–712.

Rosonina E., Kaneko S., and Manley J.L. 2006. Terminating the transcript: Breaking up is hard to do. *Genes Dev.* **20:** 1050–1056.

RNA聚合酶I和III

Grummt I. 2003. Life on a planet of its own: Regulation of RNA polymerase I. *Genes Dev.* **17:** 1691–1702.

Schramm L. and Hernandez N. 2002. Recruitment of RNA polymerase III to its target promoters. *Genes Dev.* **16:** 2593–2620.

White R.J. 2005. RNA polymerase I and III, growth control and cancer. *Nat. Rev. Mol. Cell. Biol.* **6:** 69–78.

第13章 RNA 剪接

本章概要

•

RNA 剪接的化学基础

•

剪接体

•

剪接过程

•

替换剪接

•

外显子改组（Exon Shuffling）

•

RNA 编辑

•

mRNA 转运

一个基因的编码区实际上是一系列的三联体密码子，它们决定了多肽产物的氨基酸线性顺序。迄今为止，我们一直默认编码序列是连续的，即多肽链上一个氨基酸的密码子与下一个氨基酸的密码子是直接相邻的。对绝大多数细菌及其噬菌体而言的确如此。但真核基因却很少这样，其基因的编码区被非编码区有规律地间隔开了。

所以许多真核基因是嵌合的：由一段段编码区组成，它们被另外一段段非编码区隔开了。编码区称为外显子（exon），它们之间的非编码区称为内含子（intron）。由于外显子和内含子交替出现，携带内含子的基因通常可称作“碎片”或“割裂”基因。从技术上来说，术语外显子是指在成熟 RNA 中保留的任何序列片段，不管这些序列片段是否为编码序列。非编码的外显子包括 mRNA 的 5′和 3′端的非翻译序列（第 12 章）；剪接后稳定的非编码 RNA 全长序列如 X 染色体失活调节因子 *Xist*（第 18 章）；以及有功能的 RNA 片段如在第 18 章中提到的小分子 RNA（micro-RNA）。

图 13-1 显示了一个典型的真核基因，3个内含子把基因编码区分成了4个外显子。基因的外显子数量相差悬殊，从只有一个内含子的大多数带内含子的酵母基因（和少数人类基因），再到有 50 个内含子的鸡 *proα2* 胶原蛋白基因，再到多达 363 个内含子的人 *Titin* 基因。图 13-2 显示的是一系列物种中每个基因所含内含子的平均个数，这些物种到目前为止都已测序。从图中可以清楚地看到从单细胞真核生物如酵母，到稍高等的生物如线虫和果蝇，一直到人，内含子平均数持续增加。

外显子和内含子的大小也变化不定，事实上内含子一般比它们所分隔的外显子要长得多。例如，典型外显子的长度在 150nt 左右，内含子虽说也有短的，却可长达 800kb。哺乳动物的二氢叶酸还原酶基因长达 31kb 以上，但只有 6 个分散的外显子，对应的 mRNA 长度为 2kb，因此，该基因编码区不足其全长的 10%。

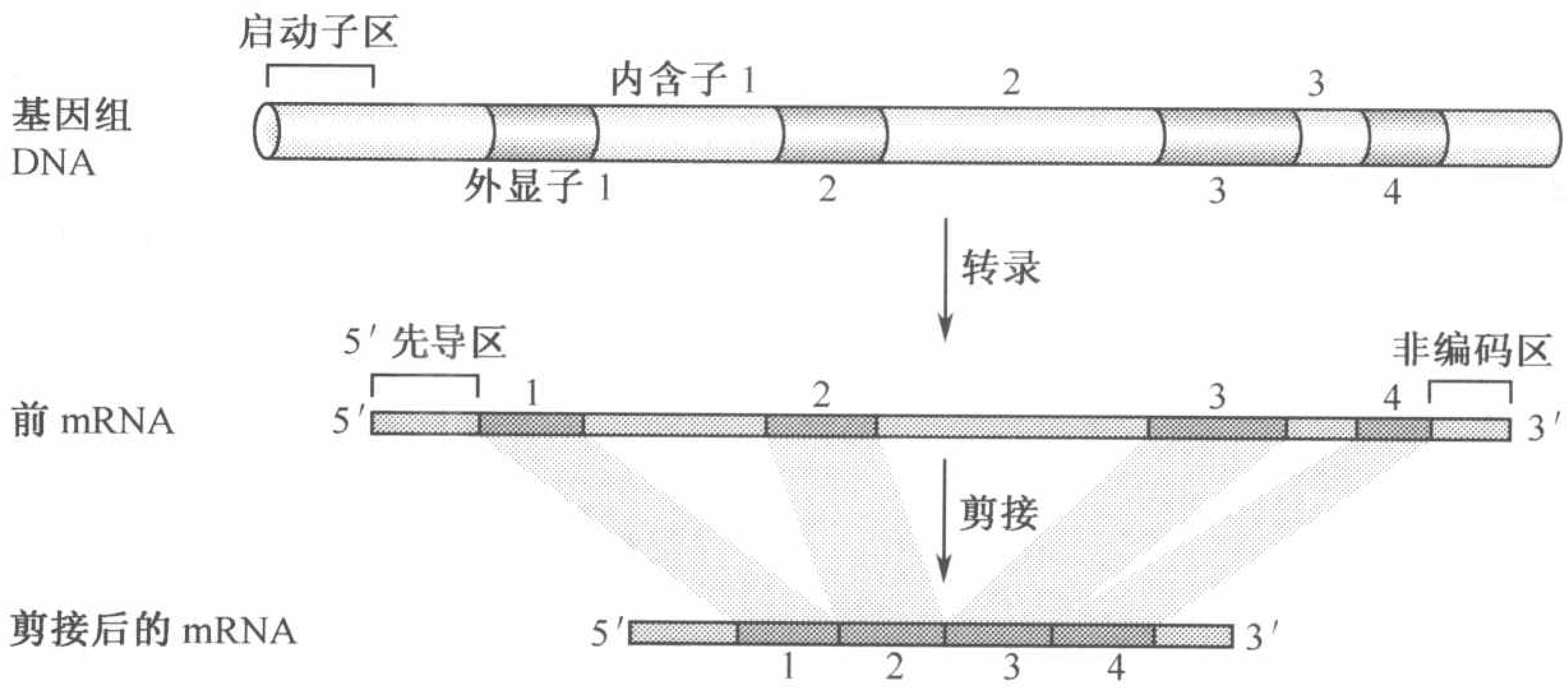


图 13-1　典型的真核基因。图中所示的基因有 4 个外显子，被 3 个内含子分开。如中间一行所示，转录从启动子区开始，产生了包括全部外显子和内含子的 pre-mRNA。剪接过程除去了内含子，并将外显子连接起来，得到了成熟的 mRNA；成熟的 mRNA 经过进一步加工（多聚腺苷酸化，第 12 章），然后转运出细胞核，就可以翻译得到一个蛋白质产物。从技术上来说，5′先导区和 3′非编码区也是外显子，因为它们保留在成熟的 mRNA 中，此处用紫色表示它们是非编码的外显子状态。

与没有内含子的原核基因一样，真核“割裂”基因也整个地转录为单一的一条 RNA 链。所以 一个典型的真核基因，其初始转录产物不仅包括外显子，也包括内含子，这显示在图 13-1 的中间一行。因为内含子数量和长度都可以非常大，初始转录产物（即前 mRNA，pre-mRNA ）实际上可以很长。人的抗肌营养不良蛋白质基因是一个极限例子，要全部转录该基因，RNA 聚合酶就得横跨 2 400kb 的 DNA 序列。（假设转录速度为 40bp/s，不难看出，要获得该基因 的一条转录体，就得花费漫长的 17h。）

虽然基因结构如此离奇，细胞的蛋白质合成机器（第 14 章）只会翻译密码子连续排列的 mRNA，它无法识别和跳过这一段段非编码序列。所以“割裂”基因的初始转录物在翻译成蛋白质之前，必须先除去其内含子。

从 pre-mRNA 中除去内含子的过程称为 RNA 剪接（RNA splicing)。这个过程将 pre-mRNA 转化为成熟的 RNA，并且必须绝对准确，避免在外显子连接处缺少或增多哪怕是一个核苷酸。我们将在第 14、15 章中看到，mRNA 的三联体密码子是按蛋白质编码序列的第一个密码子所确定的固定阅读框来翻译的。所以如果剪接不够准确，比方说在两个外显子接合处少了或者多了一个核苷酸，必将导致外显子偏离正确的阅读框，下游的密码子就会组合错误，从而把错误的氨基酸掺入到蛋白质中。

有些 pre-mRNA 存在不止一种剪接方式，从而产生不同的 mRNA。例如，可以除去不同组合的内含子，这称为可变剪接（alternative splicing)。通过这种方式，一个基因可以产生多个多肽产物。可变剪接的产物称为亚型。据估计，人类基因组中有 60% 的基因进行可变剪接，以使每个基因产生多个蛋白质产物。

通过可变剪接，从某个基因得到不同产物的数量可以从 2 种到数百甚至数千种。例如，编码神经元细胞 K^+ 通道的大鼠 *Slo* 基因，其编码产物可达 500 种之多；后面我们将会看到的，有一个果蝇基因通过可变剪接可编码多达 38 000 种产物。

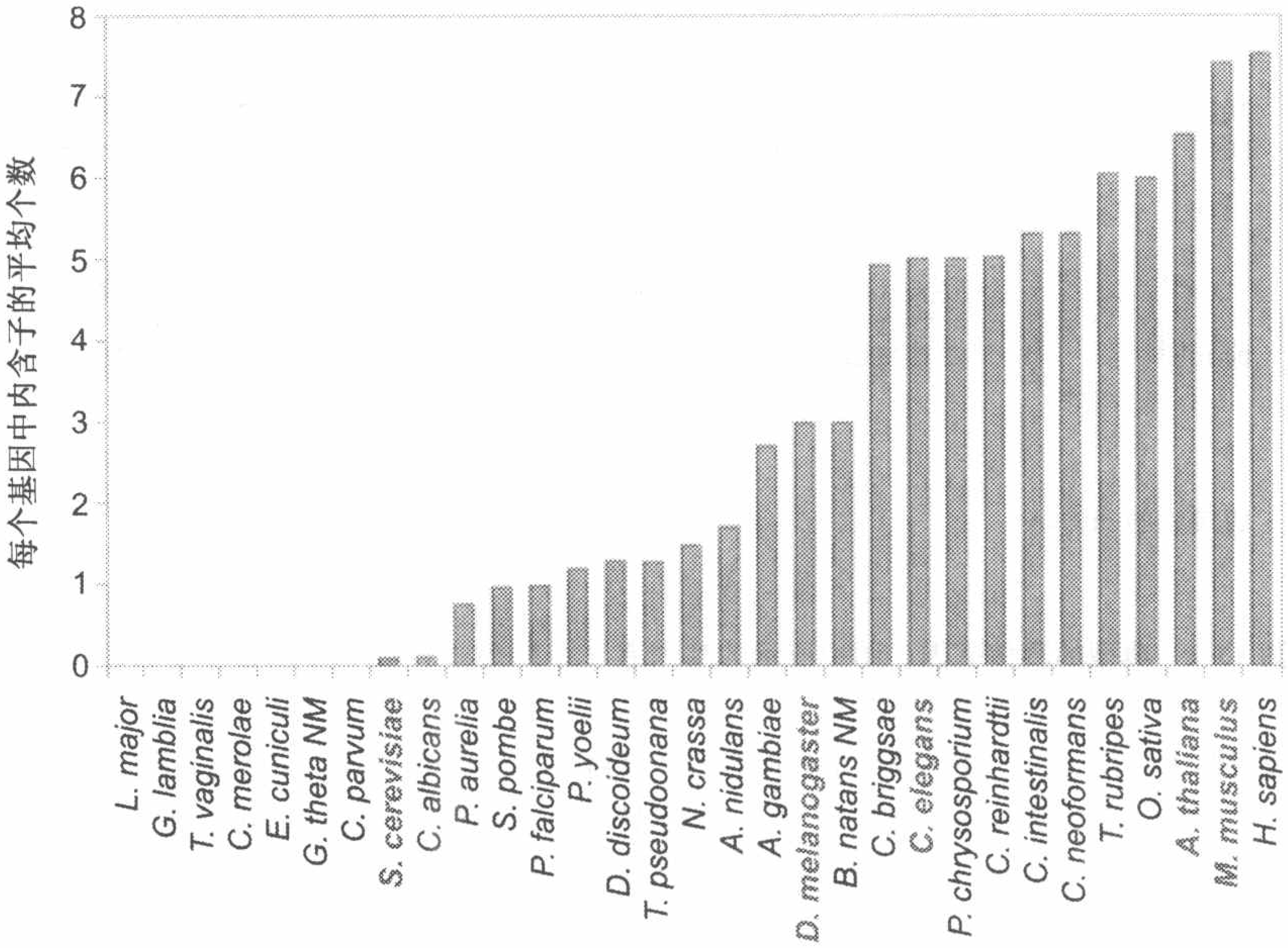


图 13-2 不同真核生物物种每个基因内含子的个数。图示为一些真核生物物种每个基因内含子的平均数。图中用红色显示的是普通的模式生物（第 12 章）：酿酒酵母（*Saccharomyces cerevisiae*），黑腹果蝇（*Drosophila melanogaster*），秀丽新小杆线虫（*Caenorhabditis elegans*），拟南芥（*Arabidopsis thaliana*），和小鼠（*Mus musculus*）。图中显示的其他物种分别为冈比亚按蚊（*Anopheles gambiae*）；构巢曲霉（*Aspergillus nidulans*）；绿领鞭藻（*Bigelowiella natans nucleomorph*）；线虫（*Canorhabditis briggsae*）；白色念珠菌（*Candida albicans*）；莱茵衣藻（*Chlamydomonas reinhardtii*）；玻璃海鞘（*Ciona intestinalis*）；新型隐球菌（*Cryptococcus neoformans*）；隐孢子虫（*Cryptosporidium parvum*）；单细胞红藻（*Cyanidioschyzon merolae*）；盘基网柄菌（*Dictyostelium discoideum*）；兔脑原虫（*Encephalitozoon cuniculi*）；蓝氏贾第虫（*Giardia lamblia*）；海藻（*Guillardia theta nucleomorph*）；人（*Homo sapiens*）；硕大利什曼原虫（*Leishmania major*）；粗糙脉孢霉（*Neurospora crassa*）；水稻（*Oryza sativa*）；双小核草履虫（*Parameciun aurelia*）；黄孢原毛平革菌（*phanerochaete chrysosporium*）；亚性疟原虫（*Plasmodium falciparum*）；鼠约氏疟原虫（*Plasmodium yoelii*）；粟酒裂殖酵母（*Schizosaccharomyces pombe*）；红鳍东方鲀（*Takifugurubripes*）；假矮海链藻（*Thalassiosira pseudonana*）和阴道滴虫（*Trichomonas vaginalis*）。（来源：Redrawn，with permission，from Roy S. W. and Gilbert W. 2006. *Nat. Rev. Genet.* 7：212，Fig. 1. © Macmillan.）

本章我们不仅讨论 RNA 剪接的机制和调控，也探讨真核基因编码区被分隔的可能原因。我们还会介绍 RNA 编辑，另一种可以改变初始转录产物以得到不同编码产物的途径。

RNA 剪接是在研究动物腺病毒基因表达的过程中发现的，详见框 13-1。

RNA 剪接的化学基础

RNA 序列决定剪接位点

我们现在考察剪接反应的分子机制（见互动动画 13-1）。内含子和外显子如何相互

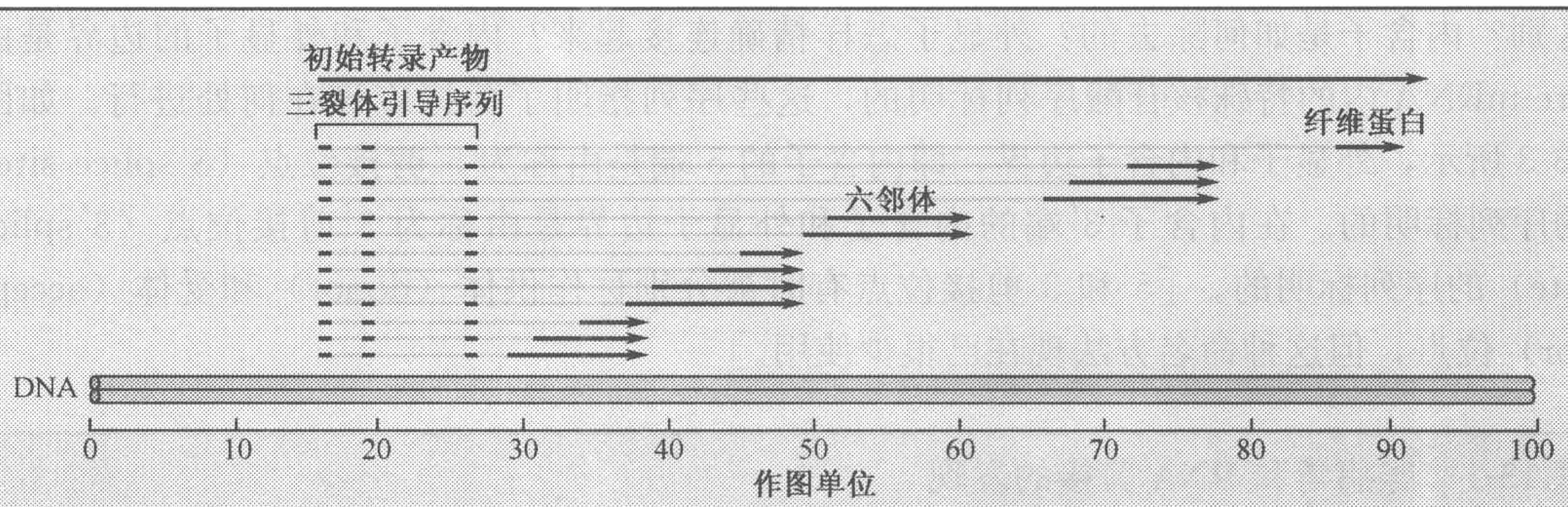
区别？内含子是如何除去的？外显子怎样精确连接起来？内含子和外显子的边界是由 pre-mRNA 内的特殊核苷酸序列标明的，这些序列指明了剪接应该在何处进行。如图 13-3 所示，外显子和内含子边界，即内含子的 5′端是由称为 5′剪接位点（5′splice site）的序列标明的。在内含子 3′端的内含子和外显子边界是由称为 3′剪接位点（3′splice site）的序列标明的。[5′和 3′剪接位点有时被分别称作供体（donor）和受体（acceptor）位点，但这种命名方法现在已很少使用。]

框 13-1 腺病毒和 RNA 剪接的发现

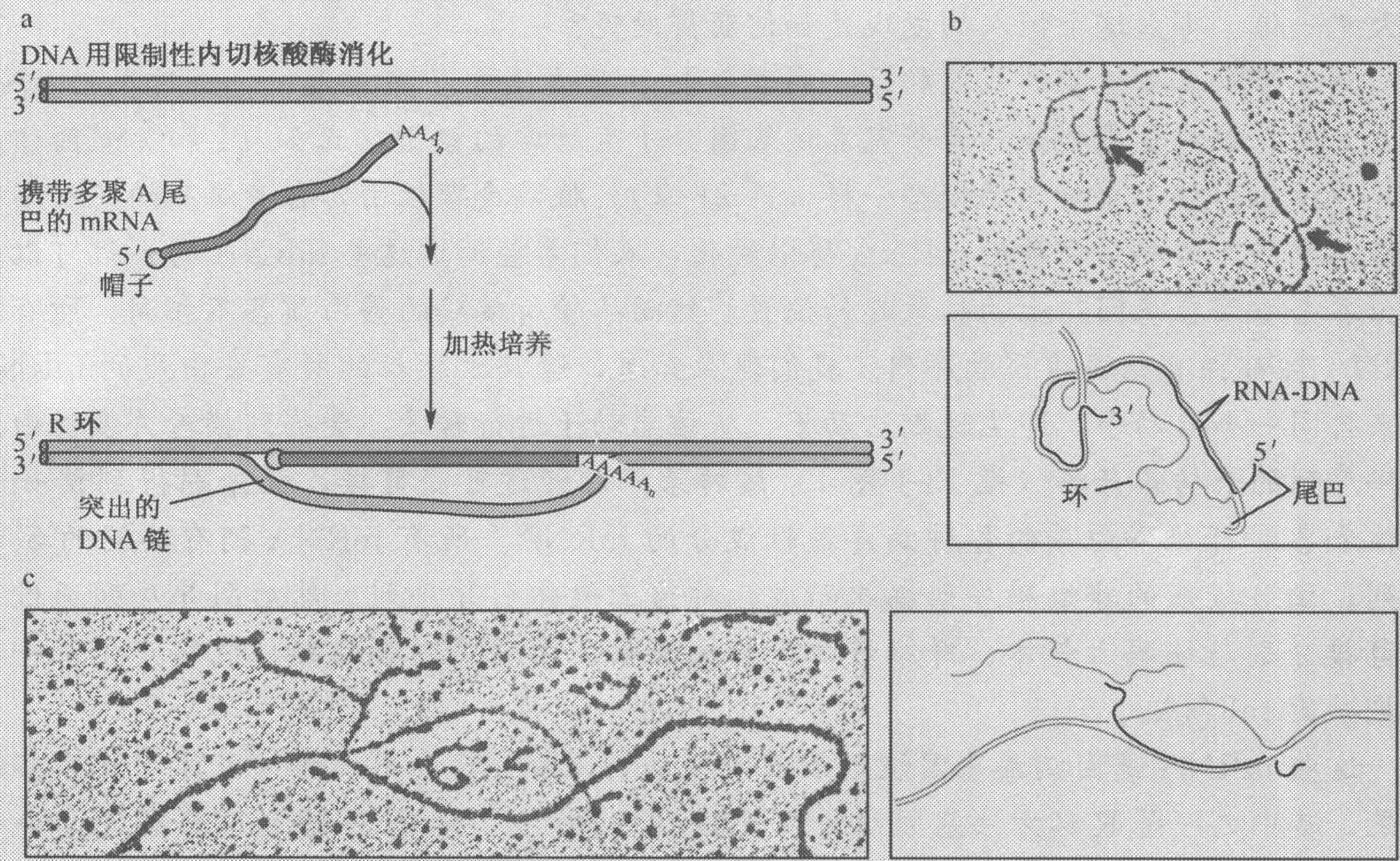
对细菌及其噬菌体的研究使人们相信：从核苷酸序列角度来讲，mRNA 是其基因的精确复制物（第 15 章）。1977 年首次发现某些（我们现在知道，实际上是许多）真核 mRNA 是从比其长得多的初始转录产物通过拼接的方式剪接得到的，当时确实让人们大吃一惊。那么这个惊人的发现是如何获得的呢？

为了了解真核生物的基因转录，科学家把注意力集中在腺病毒——一种人的 DNA 病毒上面。腺病毒最初是作为研究真核基因分子生物学的模型，正如 T4 和 λ 噬菌体是研究原核基因分子生物学的模型一样（第 21 章）。腺病毒颗粒由其自身编码的几种不同砖白质构成，这些蛋白质的 mRNA 得以纯化，人们希望通过这些 mRNA 的 5′端可以分析出各个基因在基因组上的转录起始位置。然而尽管所编码的蛋白质各不相同，所有的 mRNA 却都拥有相同的 5′端序列。我们现在知道，腺病毒颗粒的所有蛋白质的 mRNA 都来自同一个启动子，即主晚期启动子。从该启动子进行转录，会得到横跨许多蛋白质编码序列的长转录产物（框 13-1 图 1）该转录产物经过可变剪接，就分别得到诸如六邻体和各种纤维蛋白等的各种病毒颗粒成分的 mRNA。所有 mRNA 拥有共同的 5′端序列。由 3 段短的非编码序列编织而成，称为三裂体引导序列。然后引导序列通过可变剪接分别与编码六邻体、纤维蛋白和其他毒粒蛋白的序列相连接，最终产生每一种晚期病毒 mRNA。

这些 mRNA 是由病毒基因组几个不同区域的 RNA 剪接而来的结论源于许多实验结果，其中之一是 R 环方法基因组定位（框 13-1 图 2）。当 RNA 在适当条件下与包含一段相同序列的双链 DNA 一起培养时，RNA 与其互补序列杂交，使得另一条非互补 DNA 链形成一个环（框 13-1 图 2a）。然后使用观察核酸的方法进行染色，就可在电镜下看到 R 环，因为双链 DNA 或 DNA-RNA 比单链的核酸显得要粗。用腺病毒 mRNA 进行这种实验时，发现一段 DNA 上所形成的 R 环并非完全连续。相反，mRNA 的一端或两端会从 R 环伸出来成为单链尾巴（框 13-1 图 2b），具体情况视所用的病毒 DNA 片段而定。有时其中一个尾巴会与病毒基因组其他区域的 DNA 片段杂交（框 13-1 图 2c）。显然，这些 mRNA 是复合分子，由基因组上不连续区域的互补序列连接而成。这些以及其他种类的 DNA-RNA 杂交实验最终推导出框 13-1 图 1 所示的可变剪接模式。



框 13-1 图 1 人 2-型腺病毒基因组图。本图显示了病毒晚期 mRNA 的转录模式，包括初始转录产物（上端的深绿色长箭头），分别位于 16.6、19.6 和 26.6 作图单位处的三裂体引导序列（显示为绿色短线），以及编码各种病毒晚期的 mRNA 的 DNA 序列在基因组图谱上的位置（病毒晚期 mRNA 显示为深绿色短箭头）。



框 13-1 图 2 2-型腺病毒晚期 mRNA 的 R 环定位。a. 本示意图演示了 R 环结构的形成过程。用限制酶消化得到的一段双链 DNA 与 mRNA 混合，并在 80%的甲酰胺中加热到温度刚好超过 DNA 的变性温度。退火后 mRNA 与其互补 DNA 的杂交导致另一条 DNA 链的逐出。mRNA 的多聚 A 尾巴（不是 DNA 所编码的，第 12 章）也会从杂交体的一端伸出来。b. 六邻体 mRNA 与其位于 2-型腺病毒基因组的晚期区域的互补 DNA 序列培养后所观察到的 R 环的电子显微图谱及示意图。注意 mRNA 的 3′和 5′端都伸了出来。图中 DNA 用黑色线条表示；RNA 用绿色线条表示。c. 纤维蛋白 mRNA 与腺病毒全基因组和用限制酶酶切的基因组的早期区域这两种 DNA 分子培养所观察到的 R 环的电子显微图谱及示意图。（来源：EMs courtesy of b. Chow L. T.，Gelinas R. E.，Broker T. R.，and Roberts R. J. 1997. An amazing sequence arrangement at the 5′ends of adenovirus 2 messenger RNA. Cell12：1-8，page 2. Copyright © 1997. Used with permission from Elsevier. c. Berget S. M.，Moocre C.，and Sharp P. A. 1997. Spliced segments at the 5′terminus of adenovirus late mRNA. Proc. Natl. Acad. Sei. 74：3171-3175.）

图 13-3 还显示了剪接所必需的第三个序列，称为分支点（branch point site，或称分支点序列）。分支点全部位于内含子之内，通常靠近 3′端，后面为多聚嘧啶区（Py tract），如图 13-3 所示。

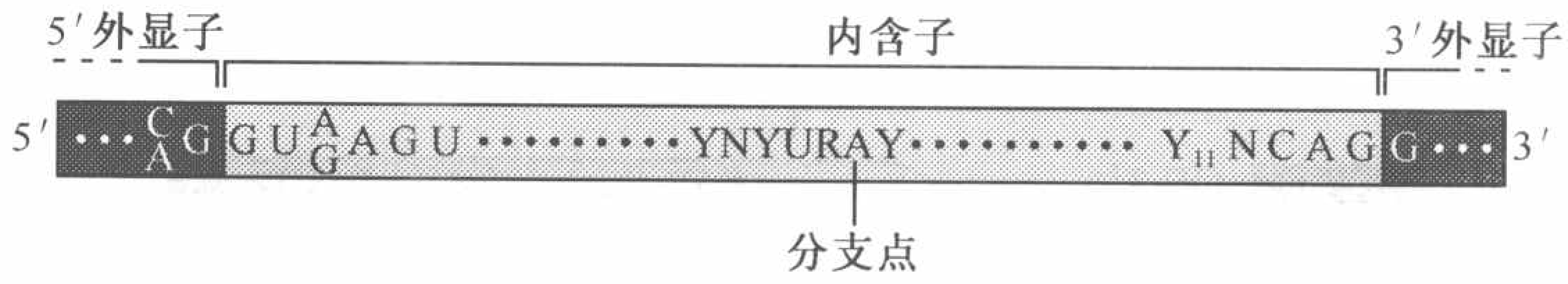


图 13-3 内含子与外显子交界处的核苷酸序列。图中显示了 5′和 3′剪接位点的共有序列和分支点的保守腺苷酸。按照共有序列标识的惯例，如果某一位置有两个核苷酸出现概率相似，就把两者都标出来。本图显示的共有序列是人类的情况。下同（除非特别声明）。

以上几个位点各自的共有序列显示在图 13-3 中。最保守的序列是 5′剪接位点的 GU 和 3′剪接位点的 AG 及分支点的 A。这些高度保守的核苷酸都位于内含子中——这也许并不奇怪，因为外显子不同于内含子，它是要用来编码特定氨基酸的，其序列必然受到限制。

当内含子两边的外显子连接时，内含子是以套索结构被除去的

让我们首先讨论剪接的化学基础。剪接是由两步连续的转酯反应（transesterification）完成的，pre-mRNA 中原来的一些磷酸二酯键断开，再形成一些新的磷酸二酯键（图13-4）。第一步转酯反应是由分支点保守腺苷酸的 2′羟基引发的。它作为亲核基团，攻击 5′剪接位点保守鸟苷酸的磷酰基团（该反应为双分子亲核取代反应，通过 5 价磷中间体完成）。

第一步反应的结果是，外显子 3′端的核糖与内含子 5′端磷酸之间的磷酸二酯键断开，而游离出来的 5′端与分支点保守腺苷酸连接。这样，除了 5′-3′骨架磷酸二酯键连接，第三个磷酸二酯键从保守腺苷酸的 2′羟基伸出来，产生一个三叉交汇点（three-way junction，保守腺苷酸因而叫做分支点）。图 13-5 示意了这种三联连接方式。

请注意，在第一次转酯反应中，5′外显子是离去基团。在第二步转酯反应中，5′外显子（准确地说，是其新游离出来的 3′羟基）反过来作为亲核基团，攻击 3′剪接位点的磷酰基团（图 13-4）。第二次转酯反应产生两个结果。首先，当然也是最重要的，是把 5′和 3′外显子连接起来了；所以第二次转酯反应才是把两个编码区真正连接起来的剪接步骤。其次，使内含子作为离去基团释放。因为内含子的 5′端已经在第一次转酯反应中与分支点腺苷酸连接，所以新释放出来的内含子形状像一个套索（lariat）。

在以上两步转酯反应中，没有增加新的化学键——断开了两个磷酸二酯键，同时又形成了两个新的磷酸二酯键。由于只是移动化学键，这个化学过程不需要消耗能量。但是，正如我们所见，在剪接反应过程中消耗了大量 ATP。这些能量不是用于化学反应，而是用来正确组装和运行剪接机器。

剪接反应的另一个要点是其方向性：如何保证剪接只沿正方向进行，即朝向图 13-4示意的产物方向进行？首先，剪接向正方向进行增加了体系的熵——一个 pre-mRNA 分子分成了两个分子，新的 mRNA 和套索状内含子。第二，切下来的内含子一旦离开剪接体就迅速降解，不能再参加逆反应。

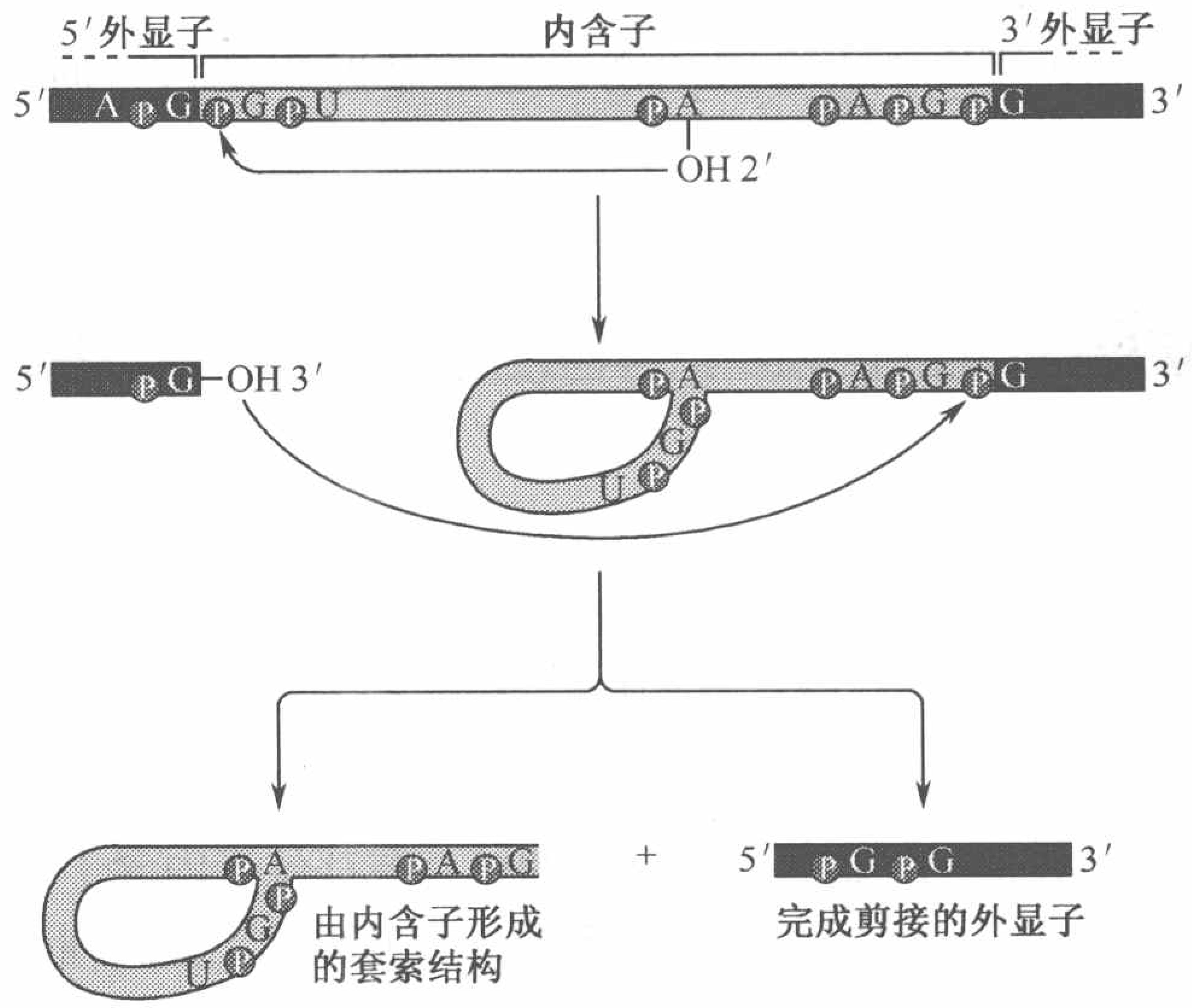


图 13-4 剪接反应。本图显示了正文中描述的剪接反应的两个步骤。在第一步中，RNA 形成环状结构，图 13-5 中有详细描述。

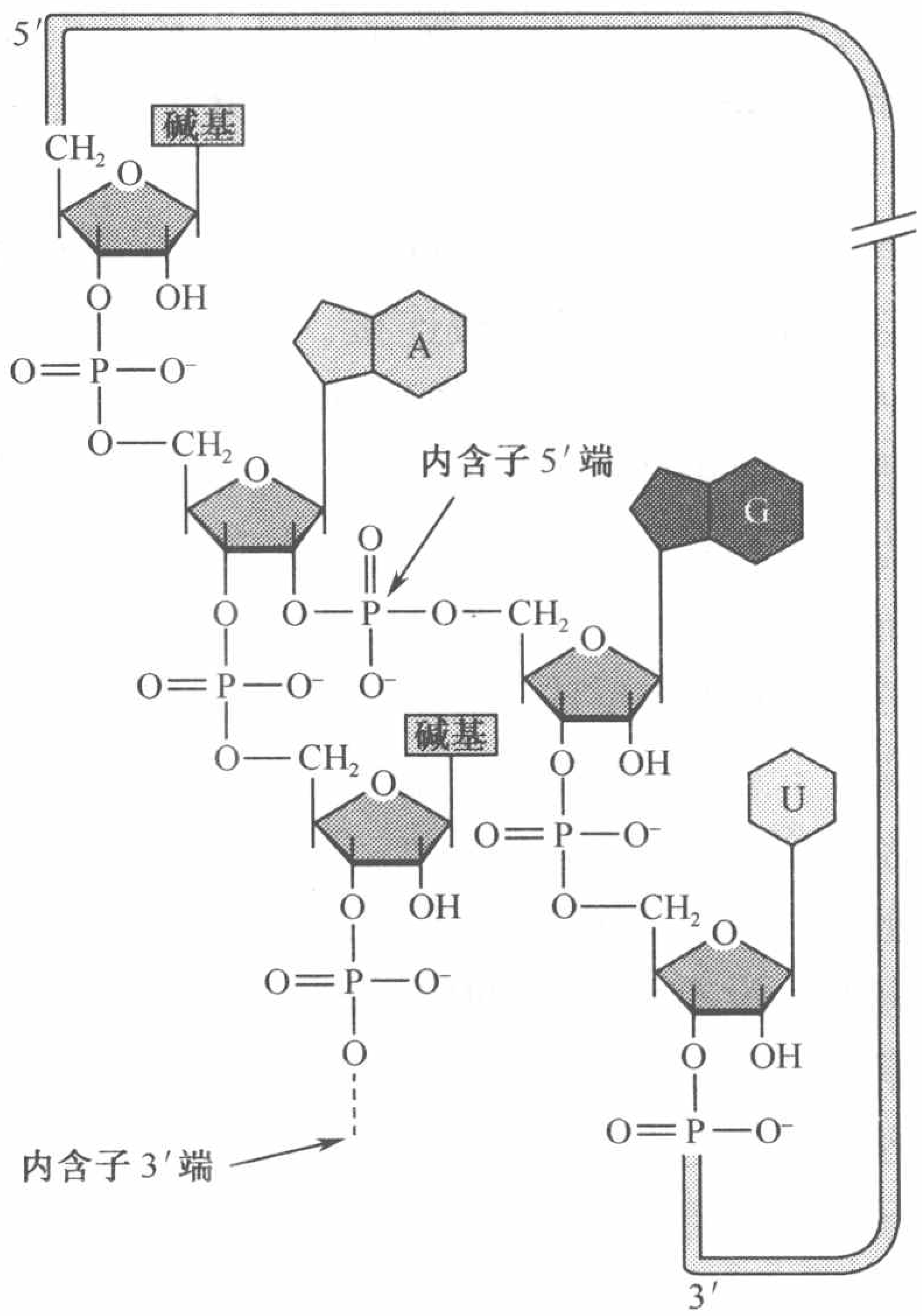


图 13-5 剪接反应形成的三叉交汇点的结构。

不同的 RNA 分子的外显子可以通过反式剪接的方式连接起来

在上述的剪接过程中，我们总是假定一个外显子的 5′剪接位点与相邻的外显子的 3′ 剪接位点直接连接。但事实上并不总是这样。在可变剪接途径中，某些外显子可以被跳过去，一个给定的外显子可以与其下游较远的外显子连接（见后文）。有时不同的 RNA 分子上的 2 个外显子能够在反式剪接（*trans*-splicing）的过程中连接到一起。尽管总体来讲反式剪接很少见，但在锥虫中反式剪接几乎见于所有的 mRNA。而在秀丽新小杆线虫中，所有的 mRNA 都进行反式剪接（以便获得一段 5′引导序列），当然许多 mRNA 也进行顺式剪接。图 13-6 显示了前述的基本剪接步骤是如何用来完成反式剪接的。

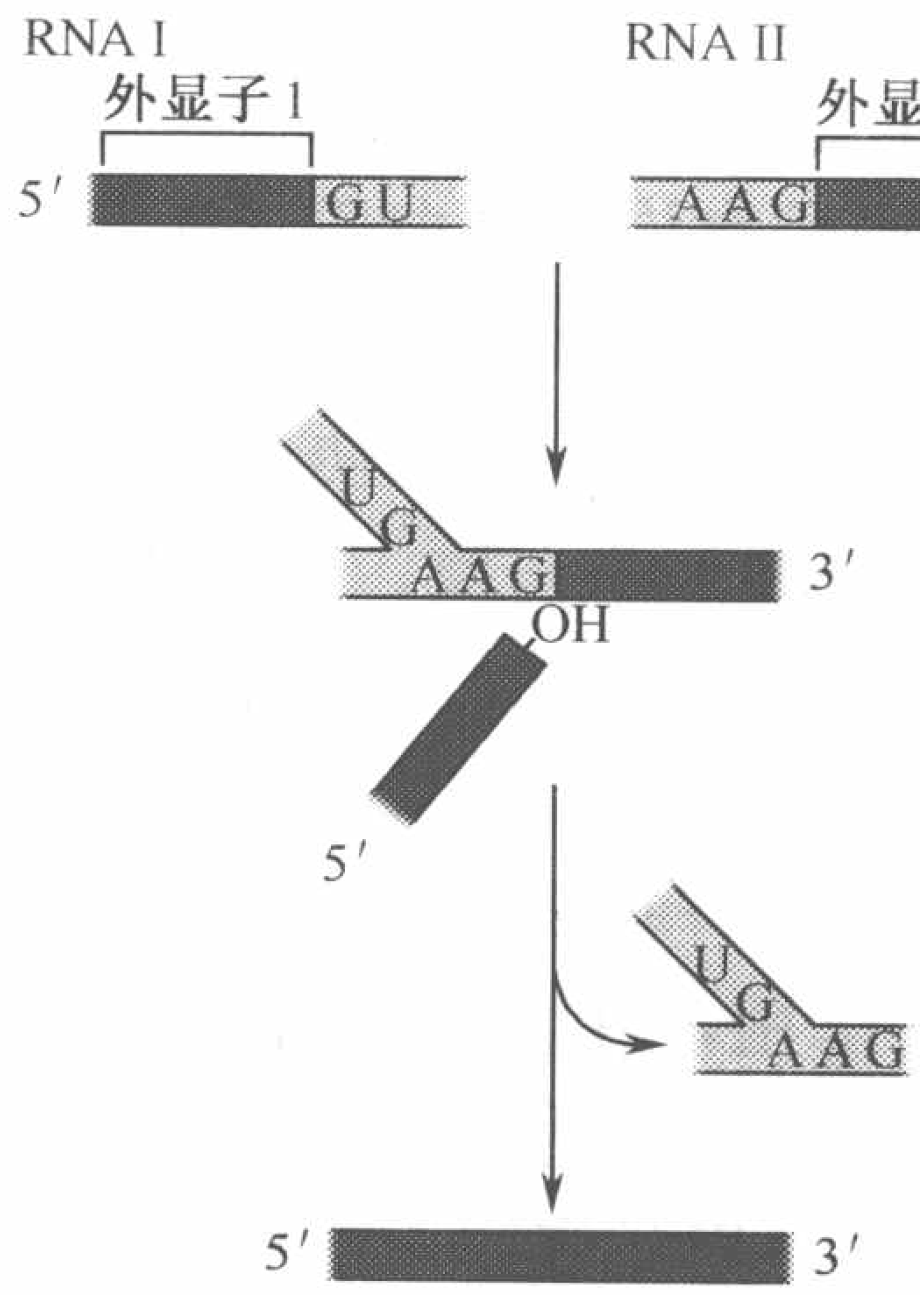


图 13-6 反式剪接。原本位于两个不同的 RNA 分子上的两个外显子，通过反式剪接可以连接进同一条 mRNA 中。反式剪接的化学过程与前述的标准剪接方式无异，剪接产物也完全一样。只是标准剪接中另一个产物是套索结构，而反式剪接中则是 Y 形分支结构。因为标准剪接是在一条 RNA 分子内形成环状结构，而反式剪接是把两条 RNA 分子连到一起。

剪接体

RNA 剪接是由一个叫做剪接体的大复合体执行的

前述的转酯反应是由称为剪接体（spliceosome）的大型分子“机器”介导的（见互动动画 13-1）。这个复合体包含约 150 种蛋白质和 5 种 RNA，大小与将 mRNA 翻译成蛋白质的核糖体差不多（第 14 章）。即使只完成一步剪接反应，剪接体也需要水解几个分子的 ATP。令人惊奇的是，剪接体的多数功能是由其 RNA 而不是蛋白质组分执行的，这再次使人回想起核糖体。因而，RNA 分子识别内含子与外显子交界处的特异序列，很可能亲自催化了剪接过程。

剪接体中的 5 种 RNA（U1、U2、U4、U5 和 U6）统称为核内小 RNA（small nuclear RNA，snRNA）。在大多数真核细胞中每种 snRNA 长 100～300nt，且与几种蛋白质形成复合物。这些 RNA-蛋白质复合物称作小核内核糖蛋白（small nuclear ribonuclear protein，snRNP）。

第 5 章中我们看到了 U1 snRNA 部分片度的晶体结构，该部分是 U1 snRNA 结合到 U1 snRNP 蛋白组成的部分。

剪接体就是由这些 snRNP 形成的巨型复合体，但是在剪接反应的不同时期具体的成分不尽相同：不同的 snRNP 在不同时间进出剪接体，每种 snRNP 执行某些专门的任务。剪接体中还有许多蛋白质不属于 snRNP，另外一些蛋白质只是与剪接体松散结合。

snRNP 在剪接中的功能有 3 个方面：识别 5′剪接位点和分支点；按需要把这两个位点集结到一起；催化（或协助催化）RNA 剪切和连接反应。要执行这些功能，所有 RNA-RNA、RNA-蛋白质、蛋白质-蛋白质相互作用都很重要。我们先考察某些 RNA-RNA 相互作用，这些相互作用发生于单个 snRNP 之内、不同的 snRNP 之间，以及 snRNP 与 pre-mRNA 之间。

例如，图 13-7a 显示 U1 snRNA 与 pre-mRNA 的 5′剪接位点之间通过互补碱基配对发生相互作用，在反应的后期，该位点又为 U6 snRNA 所识别；第二个例子见图 13-7b，分支点序列为 U2 snRNA 所识别；第三个例子见图 13-7c，显示 U2 snRNA和 U6 snRNP 间的相互作用，该作用把 5′剪接位点和分支点拉拽到一起。我们在稍后就会看到，正是这些相互作用以及其他类似的相互作用，加上由此所导致的重排，推动了剪接反应并保证其准确性。

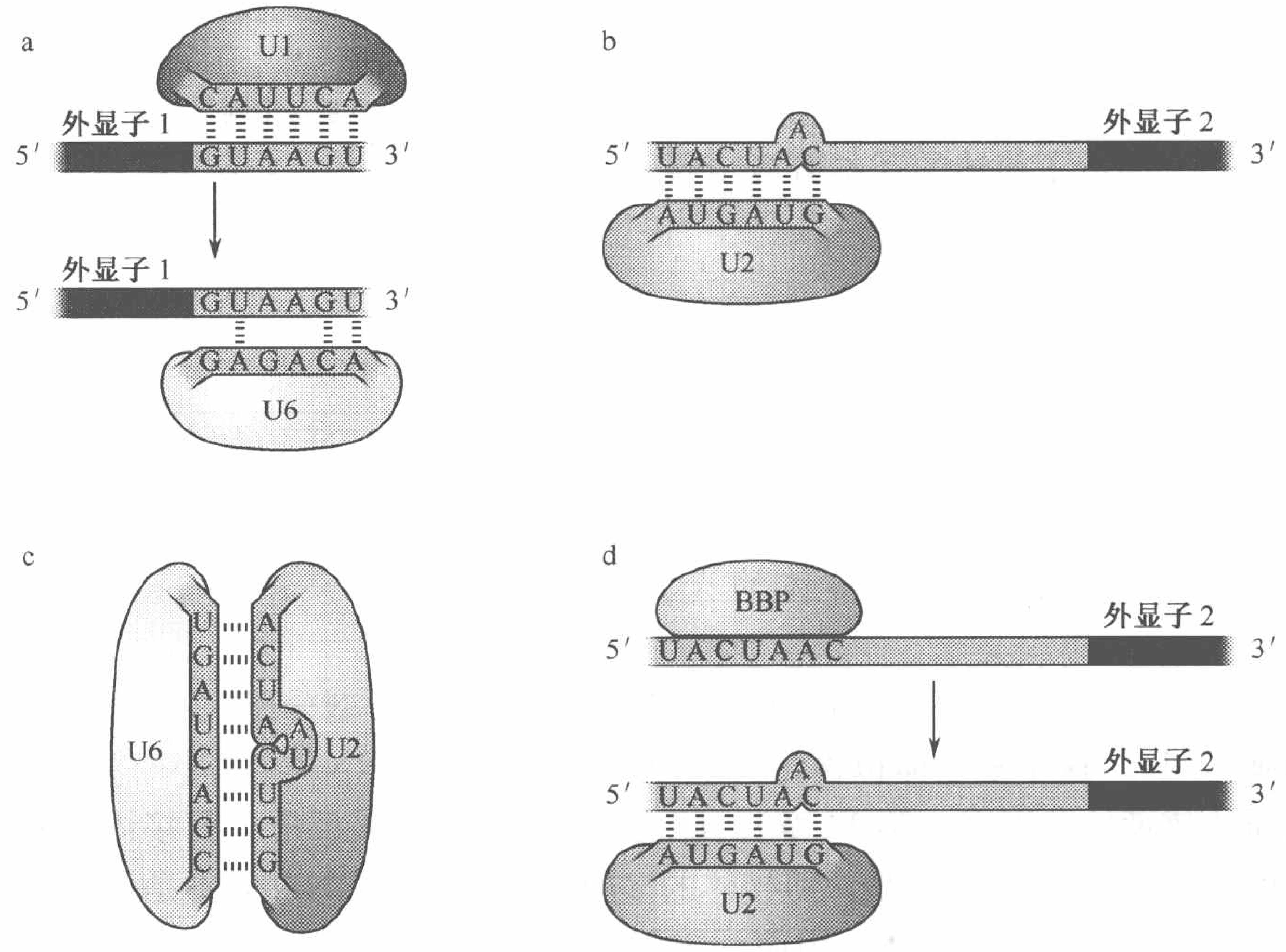


图 13-7 剪接过程中形成的某些 RNA-RNA 杂合体。 a. 不同的 snRNP 在剪接反应的不同阶段，有时识别 pre-mRNA 上的相同或交叠的序列，如 U1 和 U6 都识别 pre-mRNA 的 5′剪接位点；b. U2 snRNP识别分支点序列。c. U2 和 U6 snRNP 之间的 RNA-RNA 互补配对；d. pre-mRNA 的同一序列在不同阶段，分别为蛋白质（非 snRNP 的成分）和 snRNP 所识别。以上每一种变化都伴随着剪接体组分的来去，以及能够使剪接反应进行的结构重排。本图所示序列来自酵母。

前面提到，某些非 snRNP 也参与了剪接反应。例如，U2AF（U2 辅助因子）识别多聚嘧啶区及 3′剪接位点，并在剪接反应的起始步骤中协助分支点结合蛋白（branch-

point binding protein，BBP）结合到分支点上。如图 13-7d 所示，BBP 随后被 U2 snRNP 取代。其他参与剪接反应的蛋白质还包括某些有助于在 mRNA 上装载 snRNP 的 RNA 退火因子以及 DEAD 盒解旋酶蛋白。后者利用其 ATPase 活性解离特定的 RNA-RNA 相互作用，以便形成替代配对方式从而驱动贯穿于整个剪接过程的重排反应顺利进行。

剪接过程

剪接体内的组装、重排与催化反应：剪接过程

剪接过程的详细步骤示于图 13-8。首先，5′剪接位点被 U1 snRNP 识别（通过其 snRNA 与 pre-mRNA 之间的碱基配对，图 13-7）。U2AF 的一个亚基与多聚嘧啶区结合，另一个亚基与 3′剪接位点结合。前一个亚基与 BBP 相互作用，协助其结合到分支点上。以上形成的蛋白质和 RNA 结合体称为早期（E）复合体。

随后，在 U2AF 协助下，U2 snRNP 取代 BBP 结合到分支点，这样形成的结合物称为 A 复合体。U2 snRNA 与分支点的配对结合形成了一段双螺旋 RNA，由于分支点的腺苷酸不配对，所以被挤出来成为单碱基凸出（图 13-7b）。这样分支点腺苷酸是非配对的，可以与 5′剪接位点发生反应。

下一步是 A 复合体通过重排把 3 个剪接位点拉到一起。这个步骤是通过 U4、U6 及 U5 snRNP 加入复合体而完成的。这 3 个 snRNP 合称三联 snRNP（tri-snRNP）颗粒，其中 U4 和 U6 snRNP 通过其 RNA 组分的互补配对相结合，而 U5 snRNP 通过蛋白质-蛋白质相互作用松散结合。随着 tri-snRNP 颗粒的加入，A 复合体就转变成 B 复合体。

随后，U1 离开剪接体，其在 5′剪接位点的位置由 U6 取代。这要求 U1 snRNA与 pre-mRNA 之间的碱基配对断开，以便 U6 RNA 与同一区域配对（实际上是与一段与之重叠的序列配对，如图 13-7a 所示）。

以上就是全部组装过程。后续的重排启动了催化反应，过程如下：U4 离开剪接体，以便 U6 可以与 U2 发生作用（通过 RNA-RNA 配对，图 13-7c）。重排后的剪接体叫 C 复合体，产生了活性位点，即重排把构成活性位点的所有组分聚集到剪接体内——U2 和 U6 的 RNA 成分参与了催化反应。这个重排也保证了 RNA 底物处于合适的位置便于反应进行。令人惊奇的是，不仅活性位点主要由 RNA 构成，而且仅在剪接体组装的这个时期才形成。这种策略大概有助于降低错误剪接的概率。把剪接活性位点形成与完成剪接体早期组装过程有机地联系起来，使得活性位点只存在于真正的剪接部位。

活性位点出现后，就把 pre-mRNA 的 5′剪接位点与分支点拉到了一起，加速了第一次转酯反应。5′和 3′剪接位点的第二次转酯反应由 U5 snRNP 协助，把两个外显子连接在一起。最后的步骤是释放出 mRNA 产物及 snRNP。起初 snRNP 仍然与内含子形成的套索结合在一起，随着后者的快速降解，snRNP 又进入下一轮循环。

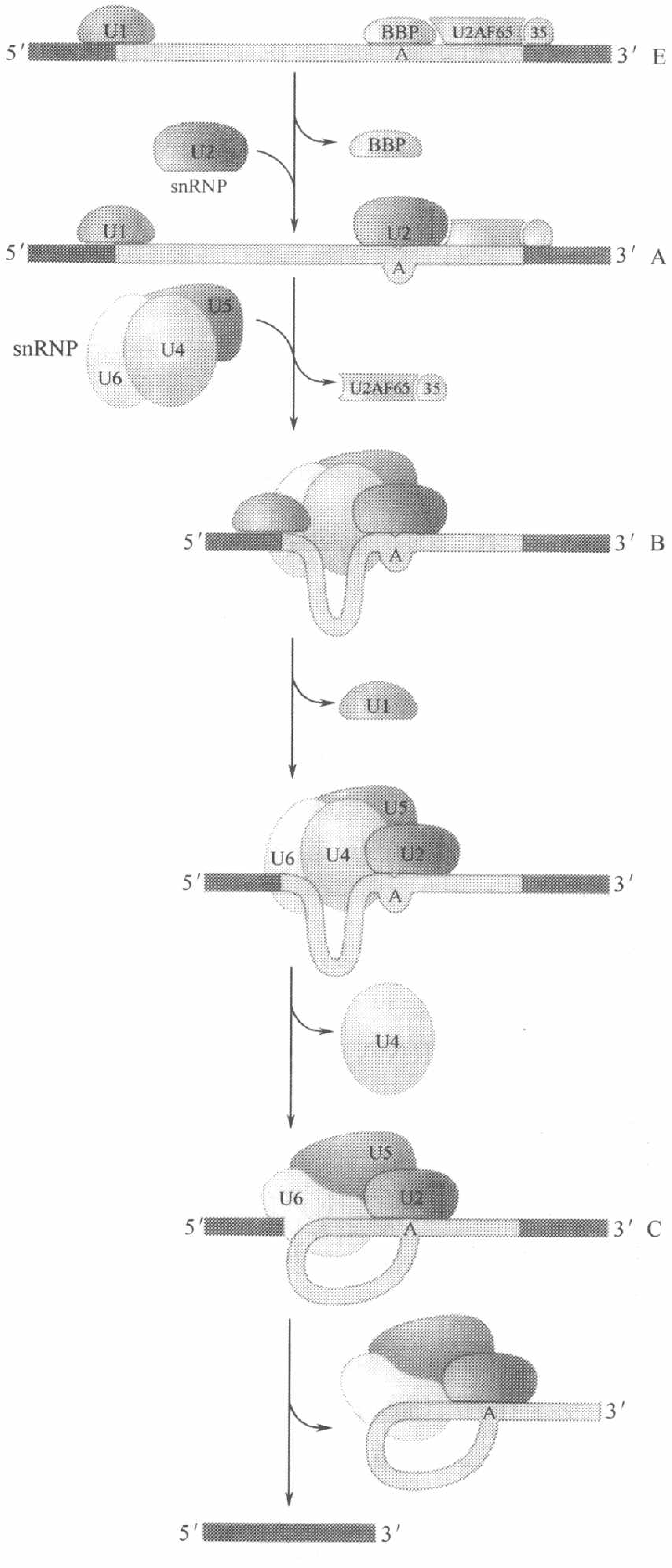


图 13-8 剪接体介导的剪接反应步骤。图中显示了剪接体的组装和作用，每一步的详细过程在正文中有叙述。在每一步骤中都有剪接体组分加入或离开。这种结构重排引起的变化是剪接反应进行所需要的。注意每个复合体的各字在右边显示。有证据表明，有些组分并非如图中表示的那一刻进入或离开；比方说它们可能还在那里但结合程度松散，而不是彻底离开。同样也不能断定某些变化过程的顺序，尤其是涉及到 U6 配对变化的两步：它先在 5′剪接位点取代 U1，还是先取代 U4 与 U2 结合？尽管有以上这些不确定的方面，但在剪接反应的不同阶段不同关键组分的参与以及剪接体的一般动力学特性还是如图所示的。

剪接机器和机制如此复杂，好像挺奇怪。剪接如何进化成这样？为什么不能简单地通过一步而不是前述的两步反应把两个外显子连接起来？要回答这个问题，让我们介绍一组内含子（与我们到目前为止所讨论的不同），它们不需要剪接体就可以把自身从 pre-mRNA 上切下来，称为自剪接内含子（self-splicing intron）。

自剪接内含子的存在表明 RNA 可以催化 RNA 剪接

细胞内已经发现的 RNA 剪接有 3 种（不包括 tRNA 加工过程，对此我们将在第 14 章中讨论），如表 13-1 所示。到目前为止我们只涉及细胞核内 pre-mRNA 剪接，它们在所有真核细胞都是由剪接体完成的。表 13-1 还显示了所谓的Ⅰ类（group Ⅰ）和Ⅱ类（group Ⅱ）自剪接内含子。自剪接的意思是前体 RNA 中的内含子自身折叠成一种特殊的构象，然后催化自身释放的化学过程（回忆一下第 6 章讨论的核酶的一般特性）。自剪接的实用定义是，在没有任何蛋白质或其他 RNA 分子存在的情况下，内含子可以在试管内将自身从 RNA 分子中剪接除去。自剪接内含子根据结构和作用机制分为两类。严格来说，自剪接内含子并不是酶（催化剂），因为它们只介导一轮 RNA 加工过程（框 13-2）。

表 13-1　RNA 剪接的 3 种类型

类型	丰度	机制	催化机器
细胞核 pre-mRNA	很常见，适用于大多数真核基因	两步转酯反应，分支点为 A	主要和次要剪接体
Ⅱ类自剪接内含子	罕见，来自某些细胞器的真核基因及原核基因	与 pre-mRNA 相同	内含子编码的核酶
Ⅰ类自剪接内含子	罕见，某些真核生物的细胞核 rRNA、细胞器基因，以及少量原核基因	两步转酯反应，分支点为 G	与Ⅱ类自剪接内含子相同

对于Ⅱ类自剪接内含子，剪接的化学过程以及所产生的 RNA 中间产物，与细胞核内 pre-mRNA 相同。如图 13-9 所示，内含子通过分支点的腺苷酸攻击 5′剪接位点（内含子 5′端与其 5′端外显子交界处）的磷酸二酯键。该反应得到如以前所示的带分支的套索结构，并在随后的第二步转酯反应中，新游离出来的外显子 3′羟基攻击 3′剪接位点，释放出套索状的内含子，并把 5′和 3′外显子连接起来。

Ⅰ类自剪接内含子释放出线形内含子，而不是一个套索结构

Ⅰ类自剪接内含子剪接机制完全不同（图 13-9c）及互动动画 13-2。它们使用游离的鸟苷或鸟苷酸，而不是分支点腺苷酸。RNA 分子结合该鸟苷或鸟苷酸，并将其 3′羟基呈递给 5′剪接位点。以前面提到的形成套索结构的同类转酯反应，把游离的鸟苷或鸟苷酸与内含子 5′端连接起来。第二步反应与以前讲述的完全相同：游离的外显子 3′端攻击 3′剪接位点。这个反应把两个外显子连接起来并释放出内含子，尽管在这里内含子是线形的，而不再是套索状。

Ⅰ类自剪接内含子比Ⅱ类小，有一个保守的二级结构（RNA 折叠见第 6 章）。该结构包括一个容纳鸟苷或鸟苷酸的结合口袋，只要这些鸟苷或鸟苷酸是以核糖的形式出现。除此之外，Ⅰ类自剪接内含子还含有一段“内在指导序列”，与 5′剪接位点的序列配对，因而确定了鸟苷亲核攻击的精确位置（框 13-2）。

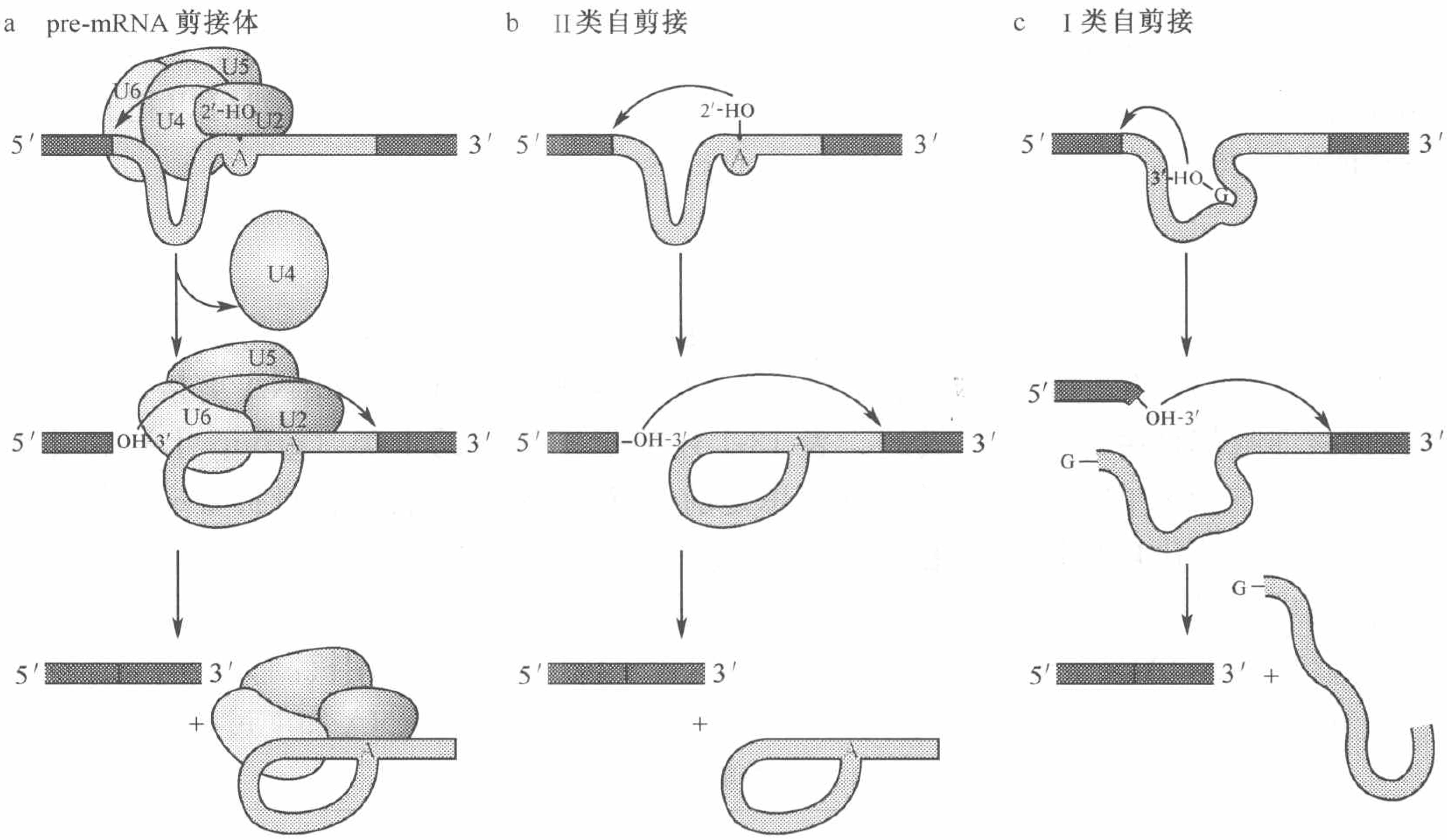
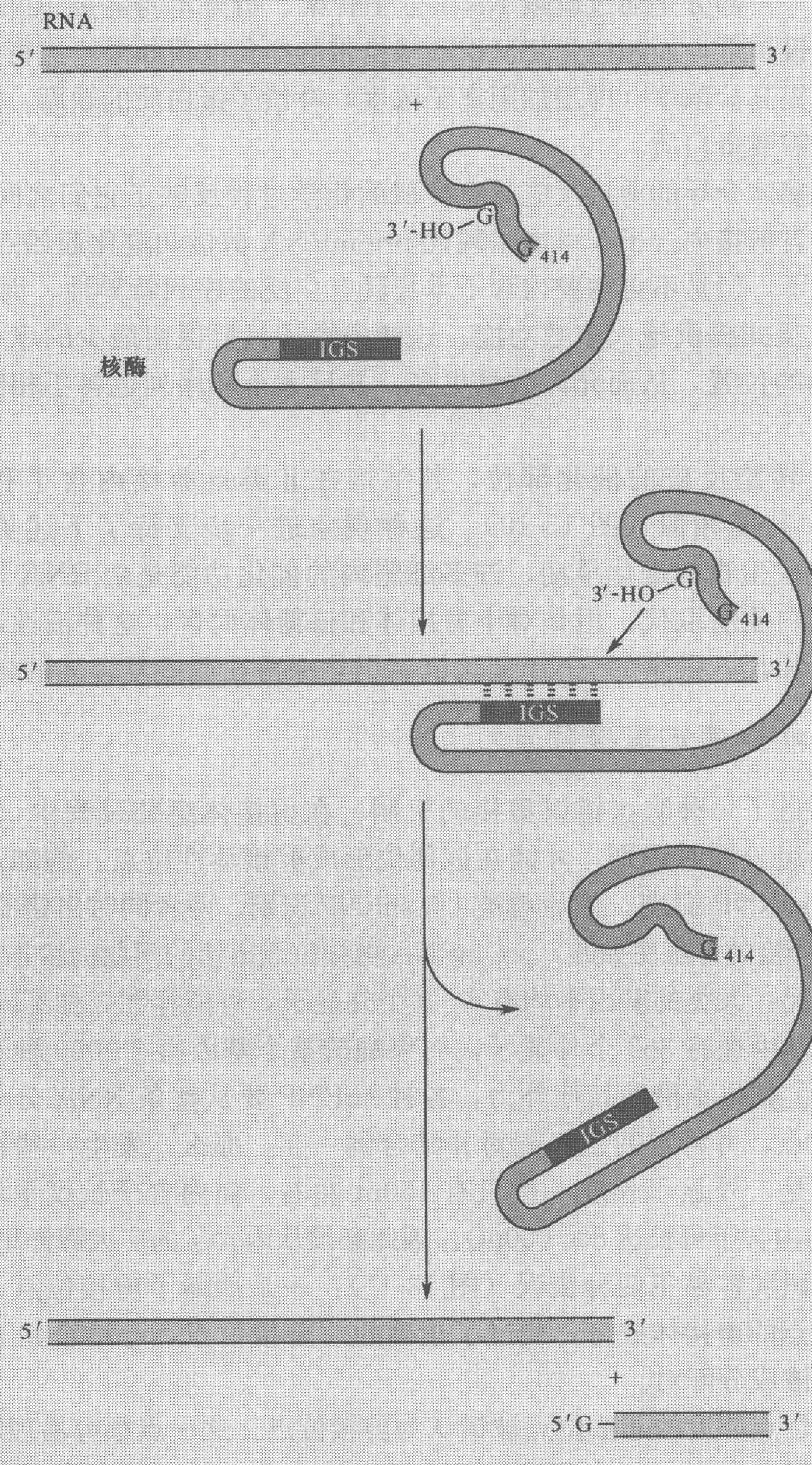


图 13-9 Ⅰ类和Ⅱ类自剪接内含子。本图比较了Ⅰ类和Ⅱ类自剪接内含子完成的剪接反应与前述的剪接体介导的剪接反应。Ⅱ类自剪接的化学过程与剪接体介导的剪接反应过程基本相同，由内含子内高度活泼的腺苷酸启动剪接过程，并形成套索状产物。对于Ⅰ类自剪接内含子，其 RNA 通过折叠形成一个鸟苷结合口袋，从而结合一个游离的鸟苷用以启动剪接过程。虽然这类内含子在体外无需蛋白质协助就能够进行自身剪接，但在体内环境下，它们通常需要蛋白质组分来激活剪接反应。（来源：Adapted with permission，from Cech T. R. 1986. *Cell* 44：207-210，Fig 1. © Elsevier）

框 13-2 Ⅰ类自剪接内含子转化为核酶

一旦Ⅰ类自剪接内含子剪接下来，其活性部位仍然是完整的。那么是什么阻止了剪接反应逆向进行呢？一个因素是细胞内高浓度的鸟苷——这非常有利于正向反应。然而除此以外，内含子本身还通过进一步的反应，有效地防止其参加逆反应。很凑巧，内含子的 3′端正好是鸟苷酸，可以进入鸟苷结合口袋。同时内含子 5′端也可以结合内在指导序列。这样就会发生第三次转酯反应，使内含子环化。与 3′端鸟苷酸形成的化学键是不稳定的，会自发进行水解。结果内含子重新线性化，但已被截短，排除了逆剪接的可能性。

如前文所述，Ⅰ类（及Ⅱ类）自剪接内含子并不是严格意义上的酶，因为其转化数只有 1。但是它们很容易通过以下方式转化为核酶（框 13-2 图 1）：上述重新线性化的内含子保留了活性部位。如果提供了游离的鸟苷或鸟苷酸，又有合适的底物正好包含与内在指导序列互补的序列，它将反复催化底物分子的切割。类似于通过将活性部位与底物分开把自剪接的斧头 RNA 改造成核酶（第 6 章），我们也可以把Ⅰ类自剪接内含子变成核酶。我们可以进一步通过改变内在指导序列来获得特定的核酶，切割我们希望的目的 RNA 分子。



框 13-2 图 1　Ⅰ类自剪接内含子可以转变成真正的核酶。

典型的自剪接内含子长 400～1000nt，与通过剪接体去除的内含子相反，其大部分序列对于剪接反应至关重要。因为自剪接内含子必须折叠成精确的立体结构才能完成相

应的化学反应。另外在体内环境下，自剪接内含子与多种蛋白质形成复合物以协助保持正确的立体结构——部分是通过遮蔽 RNA 分子骨架。折叠本身需要 RNA 骨架的某些部分与其他部分极接近，而相应骨架区磷酸根携带的净负电荷如果不进行遮蔽势必相互排斥。在体外，提高盐浓度（即增加阳离子浓度）补偿了蛋白质的缺席。这就是为什么剪接反应本身不需要蛋白质。

自剪接及剪接体介导的剪接反应具有类似的化学过程反映了它们之间的进化关系。原始的类似Ⅱ类自剪接内含子有可能是现代 pre-mRNA 剪接的进化起始点。RNA 的催化功能保留下来了，但是不再需要内含子本身具有广泛的序列特异性，而由 snRNA 及其结合的蛋白质反式提供绝大多数功能。这样内含子只要保留最少的序列以保证剪接过程发生在正确的位置，从而允许数量更多，并且大小和序列也各不相同的内含子存在。

进行第一次转酯反应的催化部位，其结构在Ⅱ类自剪接内含子和 pre-mRNA/snRNP 复合体中高度相似（图 13-10）。这种现象进一步支持了下述更广义的假设（第 6 章）：在现代生物的进化早期，许多细胞内的催化功能是由 RNA 来执行的，后来才基本上被蛋白质所取代。但是对于剪接体和核糖体而言，这种活性还没有完全被蛋白质所取代。相反，残余的 RNA 催化机制仍然保留在剪接复合体的核心部分。

剪接体怎样可靠地确定剪接位点？

我们已经知道了一种防止错误剪接的机制：在剪接体组装过程中，只有当 RNA 序列能够被多个组分同时识别，才能在该部位形成剪接活性位点。例如，5′剪接位点必须首先被 U1 snRNP 识别，然后再被 U6 snRNP 识别。两者同时出错的可能性很小，所以选择还是很严格的。虽然如此，pre-mRNA 剪接位点出错的问题仍然非常严重。

考虑以下情况：人类的基因平均有 8～9 个外显子，可能存在 3 种不同的剪接方式。但是人类有一个基因拥有 363 个外显子，而果蝇的某个基因有 38 000 种不同的剪接方式，下节将详述。如果不借助其他外力，各种 snRNP 要从整条 RNA 分子上找到正确的 5′和 3′剪接位点，并将它们正确配对并结合到一起，那么，发生一些错误就在所难免。需要指出的是，外显子长度一般只有 150nt 左右，而内含子长度平均可达 3000nt（前面提到，有的内含子可长达 800 000nt）。因此必须从内含子的广大海洋里捕捞外显子。

剪接位点的识别容易犯两种错误（图 13-11）：一是遗漏了剪接位点，例如，结合到某个 5′剪接位点的剪接体成分，跳过了正确的 3′剪接位点，与结合到下游另一个 3′剪接位点的剪接体成分配对。

二是与剪接位点邻近的非法位点被错认为剪接位点。这一点很容易理解，因为我们记得剪接位点的保守序列是相当宽松的。这样，某个 5′剪接位点的剪接体成分就可能与结合在“假”3′剪接位点上的剪接体成分错误配对（图 13-11b）。

能够增加剪接位点选择准确性的两种机制如下。其一，如第 12 章所述，在将某个基因转录为 RNA 时，RNA 复制酶Ⅱ携带了多种具有 RNA 加工功能的蛋白质（图 12-19），其中包括参与剪接的蛋白质。当在新合成的 RNA 分子上遇到 5′剪接位点时，那些蛋白质就从 RNA 复制酶Ⅱ的 C 端（剪接相关蛋白质搭乘在 RNA 复制酶Ⅱ的

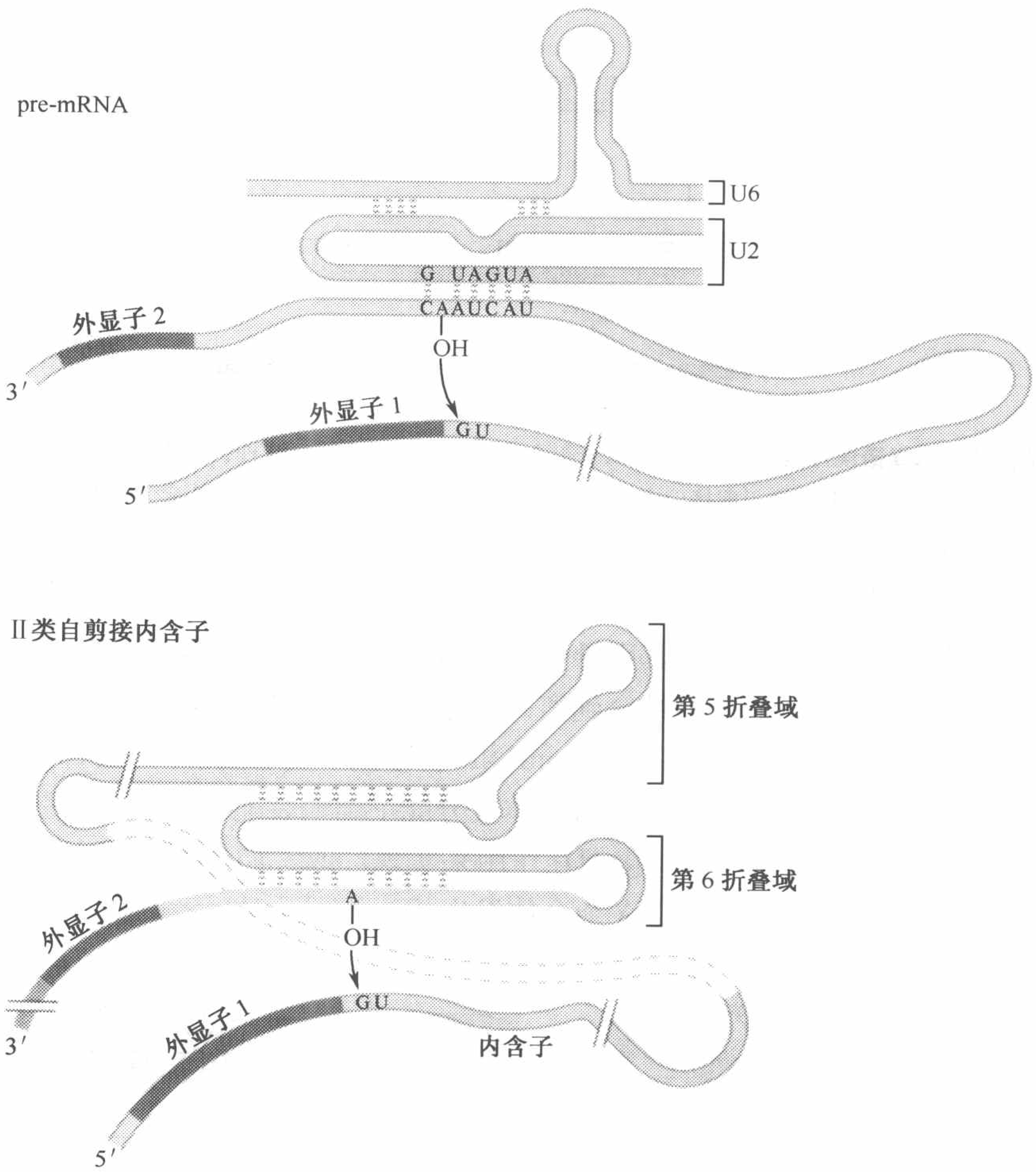


图 13-10　Ⅱ类自剪接内含子及 pre-mRNA 剪接过程中 RNA 催化区域的假设折叠方式。Ⅱ类自剪接内含子 RNA 的虚线区取代另外 4 个折叠域，本图没有描绘。

位置）转移到 RNA 分子上。剪接体成分一旦结合了 5′剪接位点，就会做好准备与那些结合即将转录出来的、下一个 3′剪接位点的剪接体成分相互作用。于是在更下游的任何竞争性 3′剪接位点被转录出来之前，剪接体成分就已经识别出了正确的 3′剪接位点。这种与转录同时进行的剪接体组装方式极大地减少了外显子遗漏的可能性。

值得注意的是，虽然多数剪接体是在基因转录的同时完成组装的——具体到每个外显子则是按其被转录的顺序，但这并不意味着内含子本身也是按相同顺序剪接。与我们熟悉的许多生物过程，如转录、复制等相反，剪接并不涉及“轨道”机制。“轨道”机制指某种复合体在基因或者信息体的一端组装，并在它向另一端移动过程中发挥作用。

其二，确保优先识别邻近外显子的剪接位点（因而更有可能是真正的剪接位点），以防止使用错误剪接位点。一种叫做富含丝氨酸和精氨酸的蛋白质（SR 蛋白）结合到外显子中的外显子剪接增强子（exonic splicing enhancer，ESE）。结合到这些位点的 SR

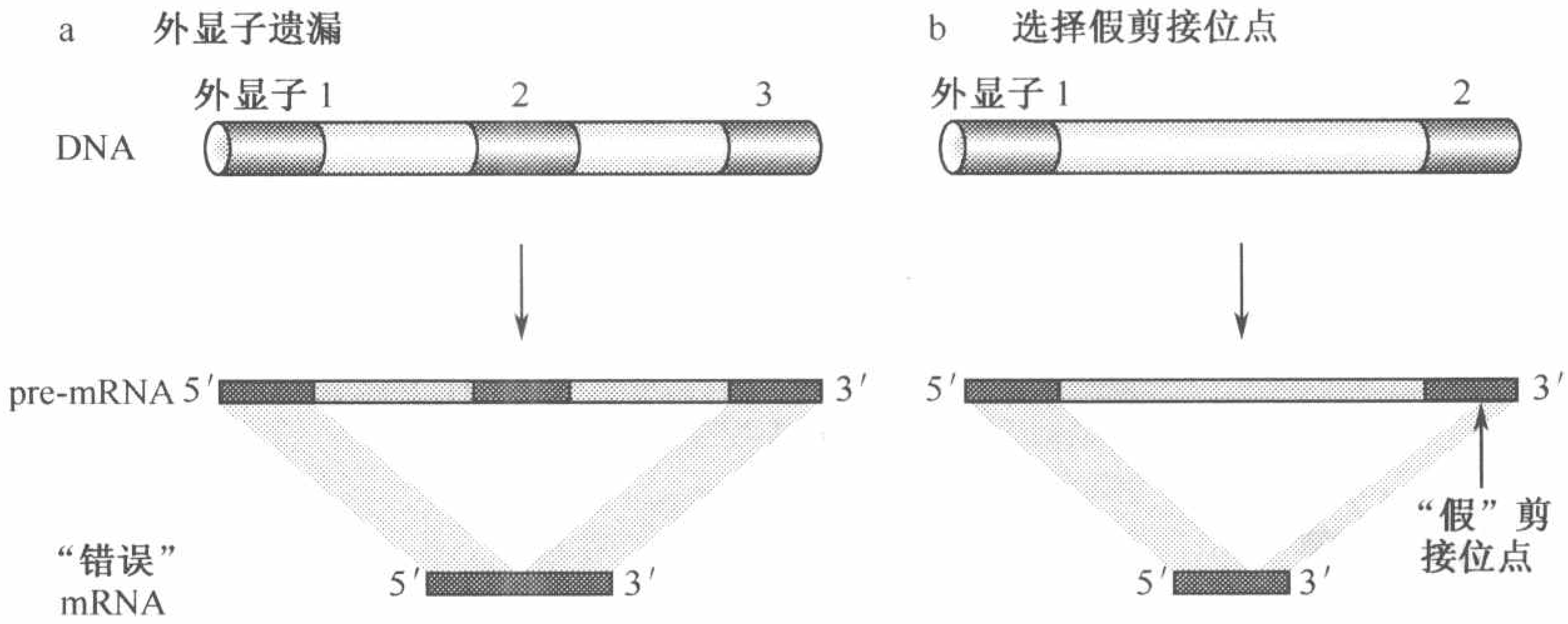


图 13-11 剪接位点选择错误可能导致的后果。a. 遗漏一个外显子的后果。这种情况发生于：结合到某个 5′剪接位点的剪接体成分，跳过正确的 3′剪接位点，与结合到下游另一个 3′位点的剪接体成分相作用。b. 图示剪接体成分识别“假”剪接位点（序列相似，但不是合法的剪接位点）所可能导致的后果。在所显示的例子中假位点位于外显子内部，使得该外显子的近 5′端被错误地当成内含子而剪接掉了。

蛋白与剪接机器的组分相作用，将其引导到附近的剪接位点。这样剪接机器就可以更有效地结合到正确的剪接位点，而不会结合到离外显子较远的错误位点。具体来说，SR 蛋白将 U2AF 蛋白引导到 3′剪接位点并将 U1 snRNP 引导到 5′剪接位点（图 13-12）。如前所述，这些因子界定了剪接位点，使其他剪接体成分能正确地装配起来。这种引导作用，或者通过 SR 蛋白和剪接体蛋白间的直接相互作用来实现，或者如最近所提出的，通过与剪接体组装和工作时形成的 RNA：RNA 杂合体发生相互作用及对后者的稳定来实现。

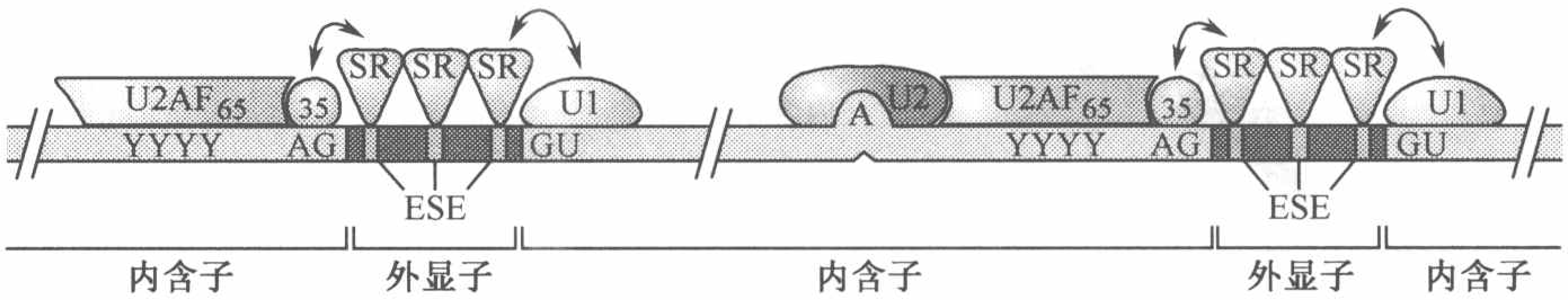


图 13-12 SR 蛋白将剪接体成分引导到 5′和 3′剪接位点。合法剪接位点由于邻近外显子，所以优先被剪接体所识别。所以 SR 蛋白结合到外显子的特定区域（外显子剪接增强子），以此为起点将 U2AF 和 U1 snRNP 分别引导到下游的 5′剪接位点和上游的 3′剪接位点。这样就在正确的位置上启动了剪接体的装配过程，也保证了剪接能够按照前面所叙述的方式进行。（来源：Adapted，with permission，from Maniatis T. and Tasic B. 2002. Nature 418：236-243. © 2002 Macmillan。）

SR 蛋白是剪接所必需的。它们不仅保证了基本剪接的准确性和高效性（上面刚刚提到），而且调节可变剪接（马上就会讨论到）。它们种类繁多，有些由生理信号调控，另外一些则不受生理调控而具有活性（组成型）。有一些只在某些特定的细胞类型中表达，调节细胞类型特异性剪接。下一节我们将讨论 SR 蛋白起作用的一些特例。

少数内含子是由另一组 snRNP 组成的另一种剪接体来剪接的

高等真核生物（包括哺乳动物、植物等）采用我们前面所提到的主要剪接体来完成绝大部分 pre-mRNA 的剪接。但是在这些生物体内（与酵母不同）有些 pre-mRNA 是由另一种丰度较低的剪接体来剪接的。这种低丰度剪接体含有某些与前述的主要剪接体

一样的组分，但还有另外一些单独的成分。低丰度剪接体中的U11和U12成分在剪接反应中的作用与主要剪接体中的U1和U2类似，但识别的序列不同。U4 snRNA和U6 snRNP在两类剪接体中都存在，虽然所使用的名称相同但实质上并不完全一样。而U5在主要剪接体和低丰度的所谓次要剪接体中是完全一样的。

次要剪接体识别一类很少见的内含子，其共有序列不同于绝大多数pre-mRNA内含子的共有序列。这种新发现的次要剪接体称为AT-AC剪接体，因为最初发现其识别的罕见内含子的5′端是AU，3′端是AC（对应于DNA则为AT和AC）。后来得知其剪接的多数内含子也具有GT-AG末端结构（与主流内含子一样），但在其他场合内含子的共有序列不同于主要剪接途径内含子的共有序列。

虽然识别的剪接位点和分支点序列不同，但是两种剪接体除去内含子的化学过程还是完全一样的（图13-13）。与这种保守的化学过程相一致，snRNP所识别的剪接位点序列差异影射snRNA互补序列的相应差异。所以，实际上保守的是剪接位点序列与snRNA碱基配对能力，并非任何一方的特定序列。

值得注意的是，AT-AC内含子可能符合前面讨论的进化图。如前所述，有人主张Ⅱ类内含子代表了内含子的最原始形式。进而，AT-AC内含子从Ⅱ类内含子进化，并最终产生主流的pre-mRNA内含子。

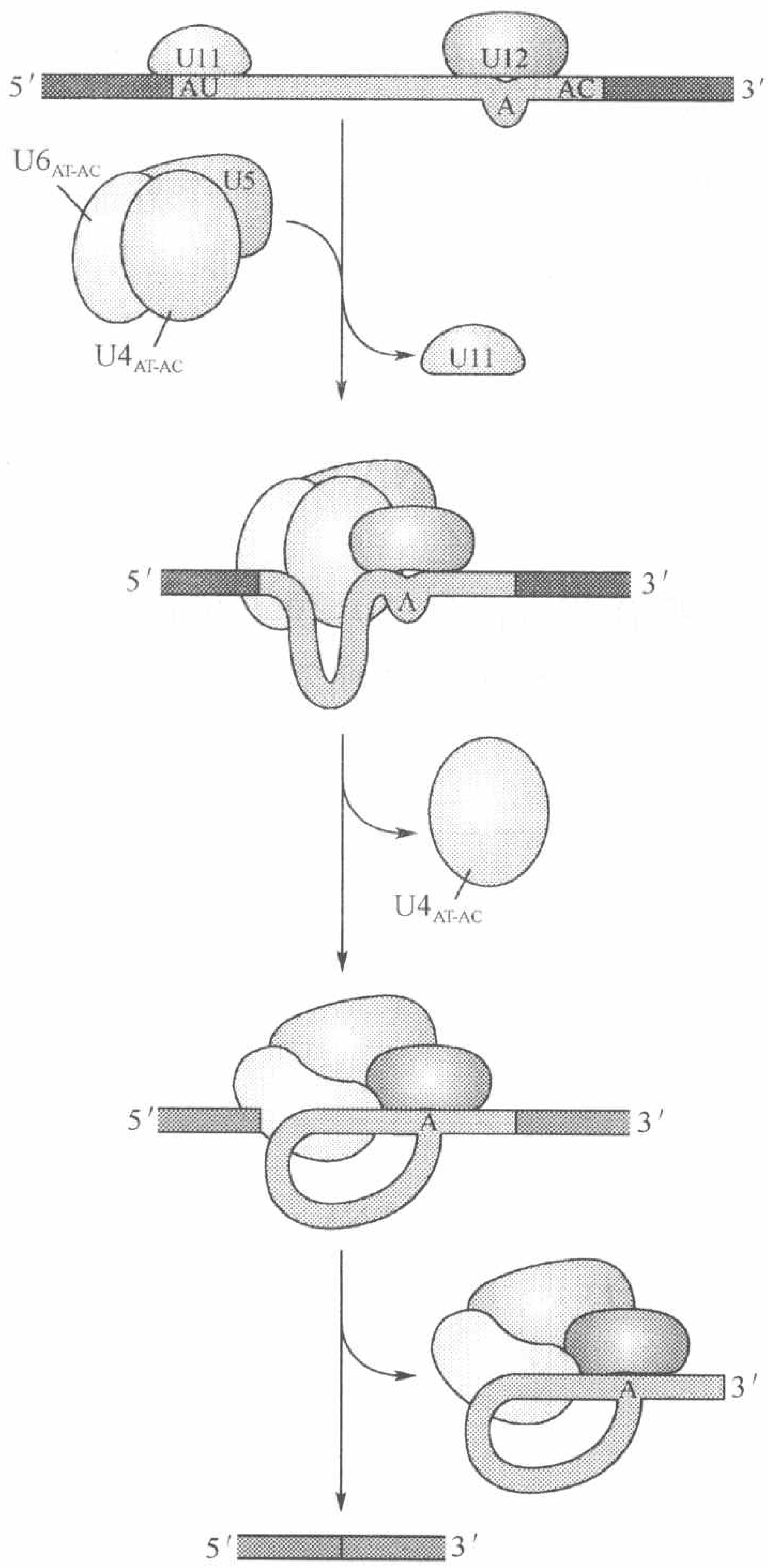


图13-13 AT-AC型（次要）剪接体催化的剪接反应。这种次要剪接体作用于少数外显子（例如，在人类可能仅为1/1000），而这类外显子有特殊的剪接位点序列。虽然如此，化学过程是一样的，剪接体成分也有相同的，其余不同的成分也是密切相关的。

替换剪接

通过替换剪接一个基因可以得到多个产物

如本章前言所述，许多高等真核生物编码的RNA可以通过替换剪接，产生两条或多条不同的mRNA，从而翻译出不同的蛋白质（或iofoms）。目前，芯片分析显示40%的果蝇基因和高达75%的人类基因需要经过可变剪接。其中很多基因仅产生两个剪接变异体，但在某些例子中，从单个基因所能得到的剪接变异体的可能数目实在惊人——

几百个（如大鼠的 *Slo* 基因）甚至几千个（如果蝇的 *DSCAM* 基因）。

一个简单的例子是哺乳动物肌肉蛋白 troponin T（肌钙蛋白 T）。图 13-14 显示该基因转录得到的一段 pre-mRNA，含有 5 个外显子。该 pre-mRNA 可以剪接成为两种不同的成熟 mRNA，各有 4 个外显子。每个 mRNA 除去一个不同的外显子，所以两者只有 3 个共同的外显子，同时各携带一个独特的外显子。但如图 13-15 所示，替换剪接可由多种方式发生。除了选择不同的外显子，外显子还可以延长（通过选择不同的下游 5′或上游 3′的剪接位点）。在其他一些情况中，外显子可以被（有意）遗漏，或内含子保留在成熟的 mRNA 中。还有一些替换剪接是基因通过不同的启动子转录所致，结果一种转录产物中包含一个 5′端外显子，而另一种产物中不含此外显子。同样，可变的多聚腺苷酸（poly-A）位点可以使 3′端外显子被延伸，或者可变 3′端外显子被某个特定基因的一些转录产物所利用。甚至还有可变反式剪接的例子（图 13-6）。

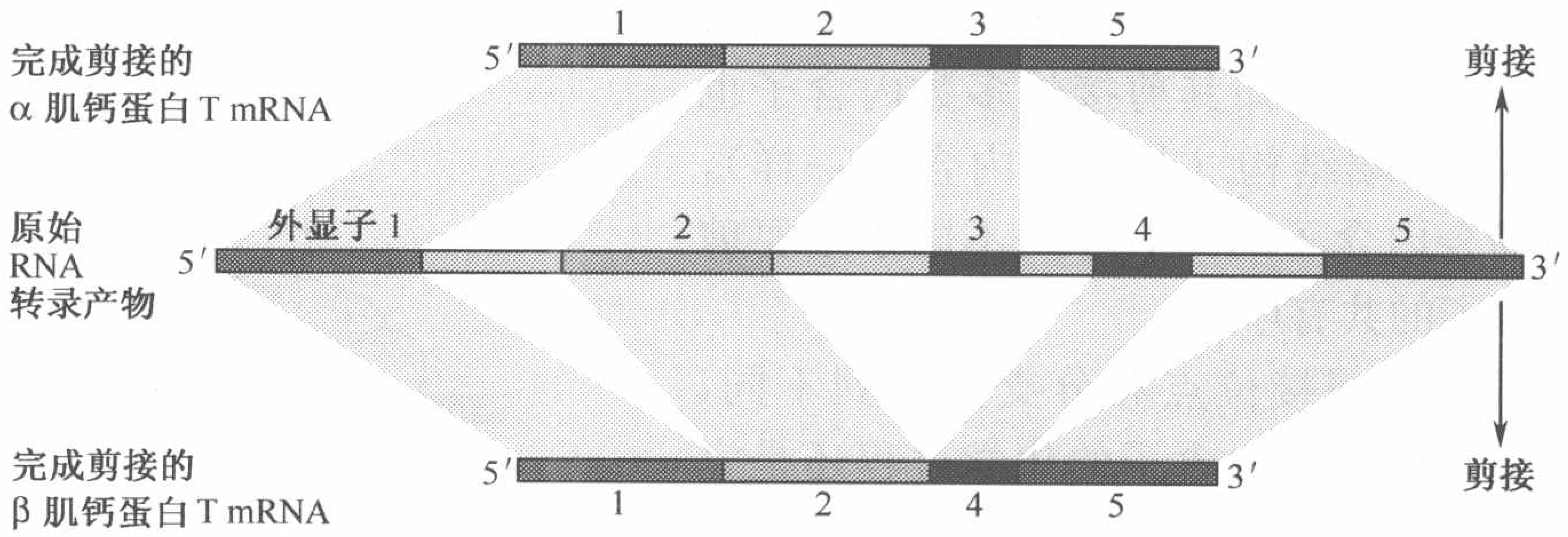


图 13-14 肌钙蛋白 T 基因的可变剪接。本图显示该基因的一段，含有 5 个外显子，剪接成为所示的两种不同形式，一种含有 1、2、4、5 外显子，另一种含有 1、2、3、5 外显子。

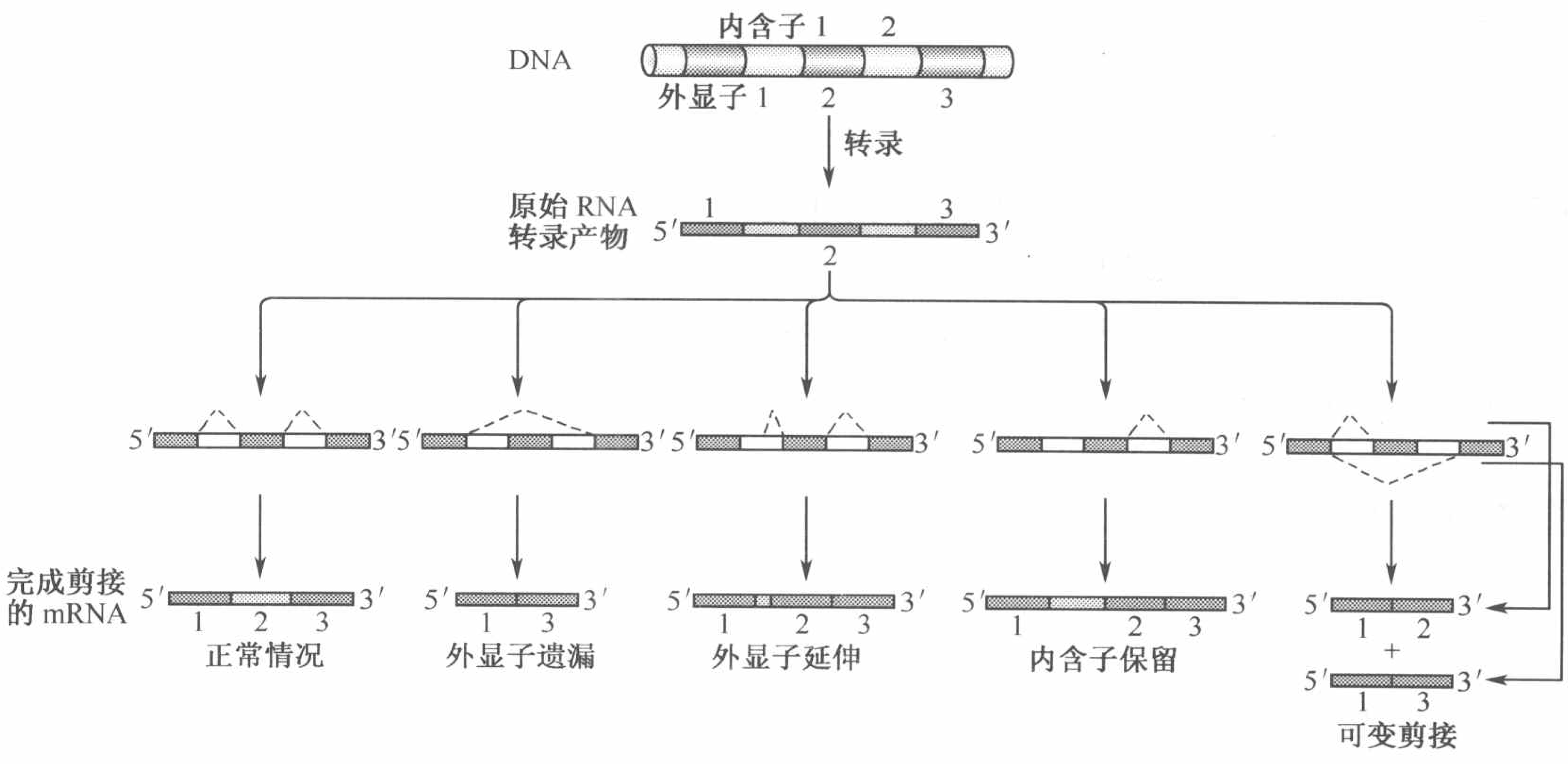


图 13-15 RNA 剪接的 5 种模式。上方表示一个有 3 个外显子的基因。中间是其转录得到的 pre-mRNA，然后通过 5 种不同的途径剪接。如果包括所有外显子，会得到含有 3 个外显子的 mRNA。如果遗漏第 2 外显子就会得到只有第 1 和第 3 外显子的 mRNA。如果第 1 外显子延伸，就会得到包括第 1 内含子的部分序列以及 3 个外显子的 mRNA。如果保留第 1 内含子，会得到包含第 1 内含子全部序列的 mRNA。而第 2 和 3 外显子用来进行可变剪接，就会产生两种 mRNA，分别包含第 1、2 外显子或第 1、3 外显子。

另一个延伸外显子剪接的例子是猴 SV40 病毒的 T 抗原（图 13-16）。T 抗原基因编码两个蛋白质：大 T 抗原（T-ag）和小 t 抗原（t-ag）。这两个蛋白质来自于同一基因的 pre-mRNA 的不同可变剪接。如图 13-16 所示，T 抗原基因有 2 个外显子，由于使用了不同的 5′剪接位点就产生了不同的成熟 mRNA。在编码大 T 抗原的 mRNA 中，外显子 1 直接与外显子 2 连接，除去了其间的内含子。反之，小 t 抗原的 mRNA 是通过使用其他的 5′剪接位点而形成的。于是，它的 mRNA 还含有部分内含子序列。（此即图 13-15 所示的外显子延伸的例子。）这个大的 mRNA 反而编码小蛋白质的原因是因为在该 mRNA 所保留的部分内含子序列中有一个符合可读框的终止密码子。

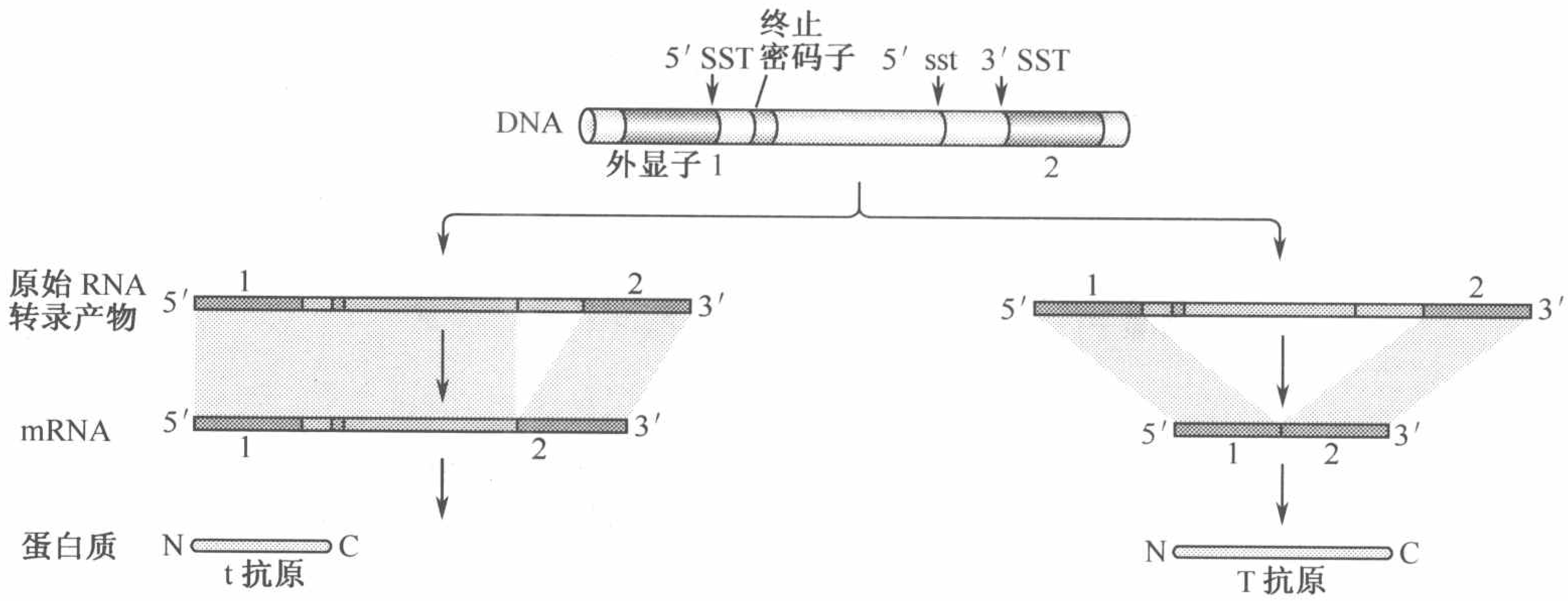


图 13-16 SV40T 抗原的可变剪接。本图显示 SV40 病毒 T 抗原 RNA 剪接过程。感染 SV40 的细胞中两种剪接方式都存在，两种抗原都产生。小 t 抗原由大的 mRNA 编码，该 mRNA 在第 2 外显子上游有一个符合可读框的终止密码子。5′SST 指产生大 T 抗原 mRNA 的 5′剪接位点；5′sst 指产生小 t 抗原 mRNA 的 5′剪接位点。3′SST 指共用的 3′剪接位点。

在感染 SV40 的细胞中，两种抗原都产生但功能不同。大 T 抗原诱导细胞转化和缩短细胞周期而小 t 抗原阻止细胞凋亡反应。两者的比例因剪接相关蛋白 SF2/ASF 的表达水平而不同。当 SF2/ASF 高表达时，驱使剪接体更多地使用最近的 5′剪接位点，从而产生更多的 t 抗原 mRNA。由于 SF2/ASF 属于 SR 蛋白，可以设想当其含量丰富时会结合到第 2 外显子的剪接增强区，然后在那儿指导组装剪接体。

在基因组范围的研究中，最常见的可变剪接形式就是完整的外显子全部包含或全部排除在成熟的 mRNA 中。这种外显子就是通常所说的盒式外显子（cassette exon）。在人约 10%的含盒式外显子的基因中，盒式外显子是成对出现的，但只有一个保留在剪接过的 mRNA 中，就像我们在 α 肌钙蛋白 T（α-troponin T）的例子中所提到的那样（图 13-14）。在这些基因中，一定存在确保外显子以互不相容（mutually exclusive）的方式进行剪接的机制。

确保互不相容性剪接的几种机制

几种机制确保了可变外显子的选择互不相容，即当一个外显子被选择留在成熟的 mRNA 上，则另一个就不被选择（如同 α 肌钙蛋白 T 的例子一样，当外显子 3 被选择

留下，则外显子 4 总是被剪切掉，反之亦然）。接下来我们讨论这几种机制，然后在下一节中讨论一个特例，它需要一种特殊的机制。

空间位阻考察一个内含子隔开的两个可变外显子。如果内含子内的剪接位点太靠近，剪接因子就不能同时结合两个位点。因此，如图 13-17 的例子所示，U 1 snRNP 结合到两个可变外显子（外显子 2 和 3）间的内含子的 5′剪接位点，阻止了 U2 snRNP 结合到同一内含子的分支点（图 13-17b）。或者，U2 snRNP 结合到分支点阻止了 U1 snRNP 结合到 5′剪接位点（图 13-17c）。α 肌钙蛋白的外显子 3 和 4 就是通过这种机制达到互不相容的。

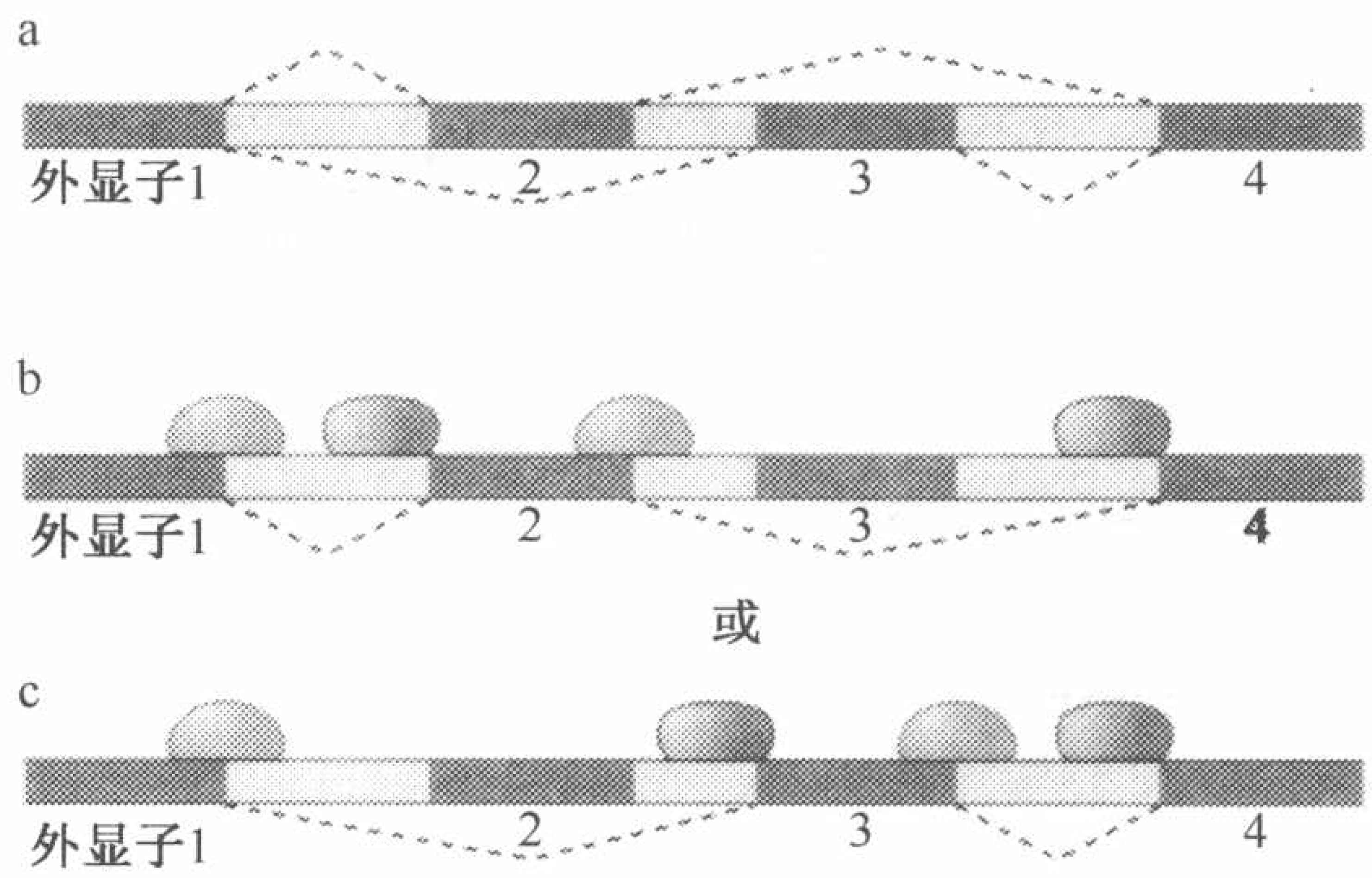


图 13-17　互不相容性剪接：空间位阻。 a. 图示可变剪接的多种可能性。b. U1 snRNP 结合到第二个内含子的 5′剪接位点阻止了 U2 snRNP 结合到该内含子的分支点；U2 结合到第三个内含子导致外显子 3 被剪切掉。c. U2 snRNP 结合到第二个内含子的分支点阻止了 U1 结合到该内含子的 5′剪接位点，此时，U1 结合到第一个内含子的 5′剪接位点导致外显子 2 被剪切掉。（Brenton Graveley 惠赠.）

这种重排可以起因于一个内含子内剪接位点的相对位置或者由于内含子就是太小不能发挥作用。在果蝇中，任何小于 59 个核苷酸的外显子均属于这一类。

主要剪接位点和次要剪接位点的联合如前所述，除了主要剪接体外，还存在另一种形式的剪接体即所谓的次要剪接体，它识别不同于前者的剪接位点。这两种剪接体都不能去除含组合位点（即一种类型的 5′剪接位点和另一种类型的 3′剪接位点）的内含子。因此，通过明智地排列这些可变剪接体所识别的 5′和 3′剪接位点，达到可变外显子的互不相容，如图 13-18a 所示。人 *JNKl* 基因就是如此。

无义介导降解与迫使剪接机器以互不相容的形式剪接不同，此机制确保只含一个或另外一个外显子（两个内含子不会同时出现或同时不出现）的 mRNA 能存活。换句话说，尽管不能保证互不相容性剪接，但这种机制的结果和互不相容性剪接一样。无义介导降解（nonsense-mediated decay，NMD）基于的事实是，同时包含两个外显子产生的 mRNA 含有提前的终止密码子（图 13-18b）。这种 mRNA 会被 NMD 破坏，细节在第 14 章中讲述（图 14-50）。

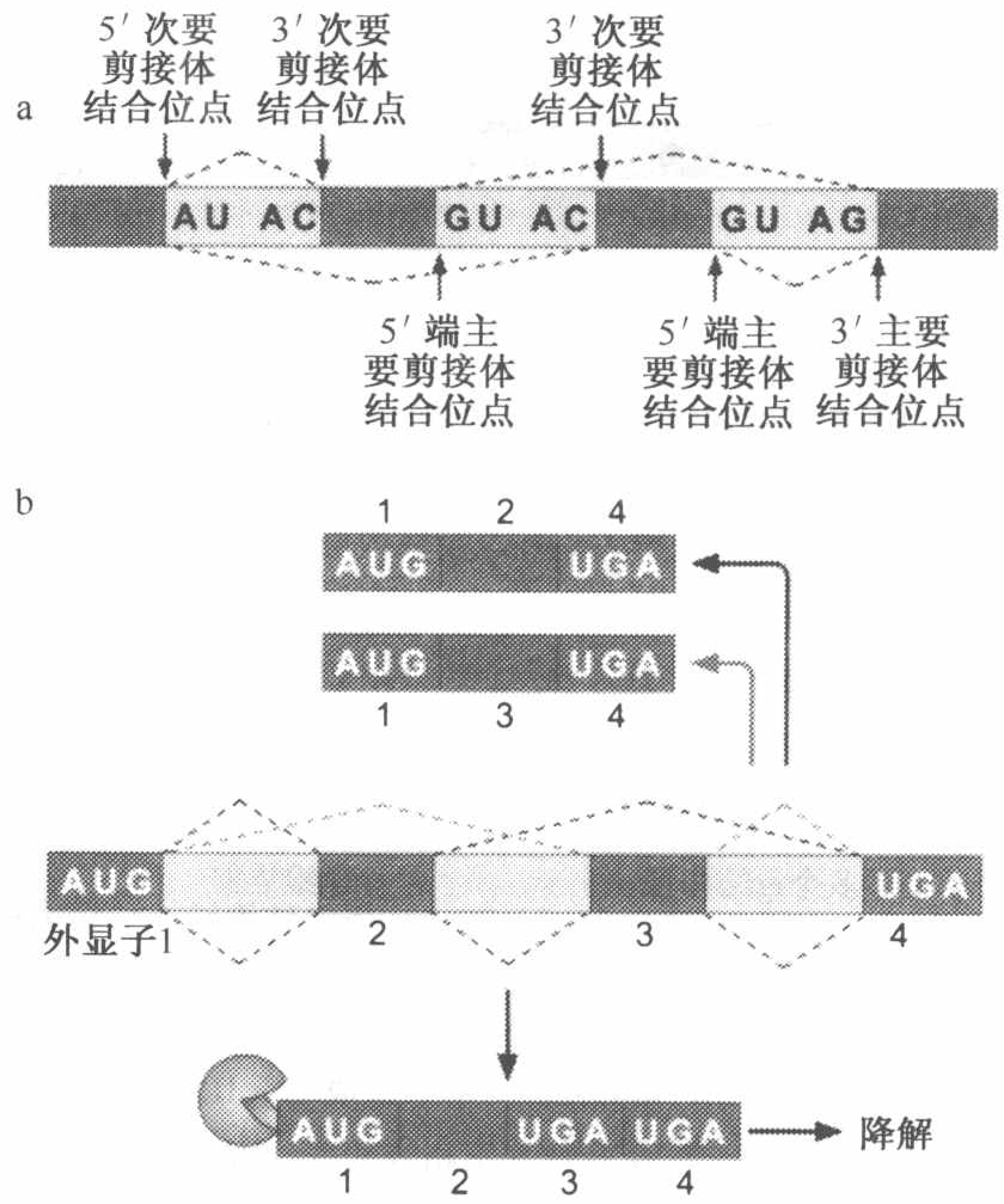


图 13-18　互不相容性剪接。a. 主要和次要剪接体识别的剪接位点。b. 无义介导降解。（来源：Brenton Graveley 绘制、惠赠.）

关于果蝇的 *Dscam* 基因的特例：大范围的互不相容性剪接

果蝇的 *Dscam*（唐氏综合症细胞黏附分子，Down syndrome cell-adhesion molecule）基因能够编码 38016 种不同的蛋白。如图 13-19 所示，来自 *Dscam* 基因的每一种可能的 mRNA 都含有 24 个外显子，其中 20 个是相同的，另外 4 个外显子（外显子 4，6，9 利 17）在 pre-mRNA 上是以多种可变形式存在的。因此，外显子 4 有 12 种可变版本，外显子 6 有 48 种，外显子 9 有 33 种，外显子 17 有 2 种。这些外显子可能的排列（12×48×33×2）产生大量可能形式的成熟 mRNA。

Dscam 基因编码的蛋白是细胞表面蛋白，属于免疫球蛋白（immunoglobulin，Ig）超家族成员。图 13-19 下部显示的是该蛋白的一般结构。此蛋白分子含有一个跨膜区域（由外显子 17 编码，因此有两种可变形式）；各种亚型中都相同的纤维连接蛋白结构域；以及 Ig 结构域，其中三个 Ig 结构域的部分片段是由高度可变的外显子 4，6 和 9 所编码的。因此，不同亚型的可变性主要来自于 Ig 的这些结构域。

在果蝇中，Dscam 蛋白执行两种完全不同的功能：既可以在大脑的神经网络模式中起作用，也可以作为天然免疫系统的一部分识别抗原。在执行神经功能时，Dscam 蛋白介导特异的细胞与细胞间相互作用。任何特定亚型的蛋白只能和与同亚型相互作用，而不能和其他亚型相互作用。人们认为这种选择性确保一个特定的轴突能够区分其他轴突，依据是否表达与其相同或者不同的亚型。因此，人们认为 Dscam 蛋白在建立大脑神经网络中起关键作用。

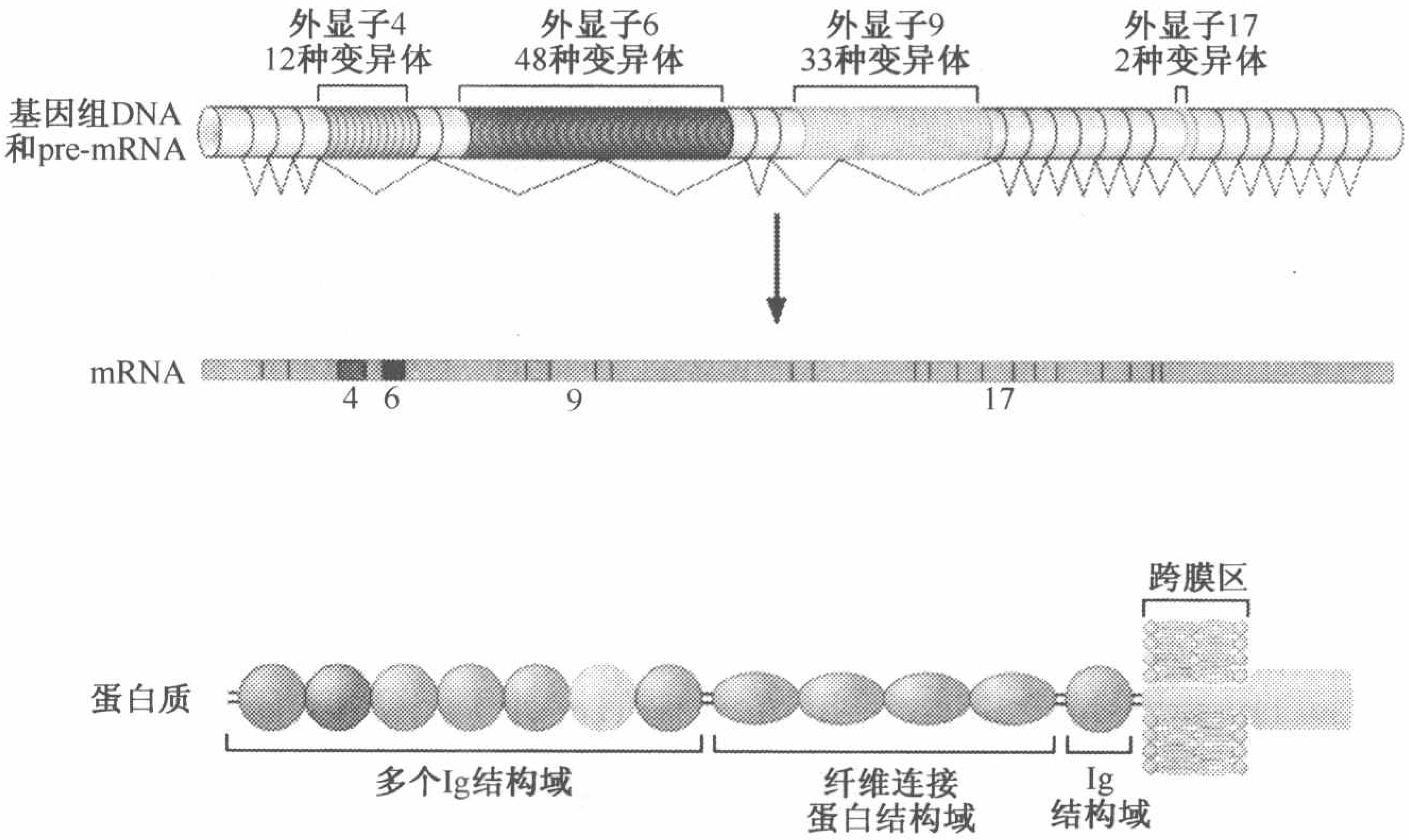


图 13-19 果蝇 *Dscam* 基因的多重外显子。 *Dscam* 基因（图的上部）长 61.2kb。经转录和剪接，产生一种或多种长度为 7.8kb，由 24 个外显子组成的 mRNA（图中显示这些 mRNA 的一般结构）。外显子 4，6，9 和 17 存在多种互不相容的选择方式。因此，每一个 mRNA 都包含一个外显子 4（12 种可能选择中的一个，用红色表示），一个外显子 6（48 种选择，用紫色表示），一个外显子 9（33 种选择，用蓝色表示），以及一个外显子 7（2 种选择，用黄色表示）。外显子 4，6，9 编码二个 Ig 结构域的一部分，用相应颜色表示；外显子 17 编码跨膜结构域。如果用到这些外显子的所有可能组合，*Dscam* 基因能产生 38016 种不同的 mRNA 和蛋白。（来源：Schmucker D. 2000. Ce//101：671，Fig. 8. © Elsevier 经许可改绘）

在免疫系统中，不同亚型识别不同的抗原，就像脊椎动物体内的抗体一样。人们认为进化压力驱动的多样性来源于这种功能的选择。

Dscam 外显子 6 的互不相容性剪接无法用任何一种标准机制来解释，而要用一种新的策略来解释

如我们前面所看到的，外显子 6 是 *Dscam* 基因中 4 个可变剪接外显子中的一个，在这个例子中，外显子 6 有 48 种可供选择的版本。这种选择规模用前几种选择性剪接机制无法解释。例如，尽管空间位阻可以解释毗邻外显子不被剪接，但不能解释其他相距更远的外显子同样不被剪接。此外，在 *Dscam* 基因上所有的剪接位点都是主要剪接体位点，所以双剪接体机制不能解释。NMD 同样无法解释外显子 6 的互不相容剪接：即使由于包括 0 个、2 个或 3 个外显子而导致移码，一个含有 4 个外显子的 mRNA 与只有一个外显子的 mRNA 具有相同的读码框。含 7 个外显子的 mRNA 也是如此，依此类推。

那么细胞怎么保证在 mRNA 中只有一个外显子 6 的变异体呢？新机制依赖 pre-mRNA 中可变的 RNA：RNA 碱基对结构。每种可变结构确保仅有一个外显子 6 的变异体在任何时候都得到保护，不受普遍存在的剪接抑制所抑制。我们现在考察这种机制是怎么工作的，以及通过果蝇 *Dscam* 基因和其他昆虫物种中相应基因的序列分析，这

种机制是怎么被发现的?

图 13-20 显示的是这种机制的基本模型。该模型有两类保守序列元件。第一个元件停泊位点(docking site)位于外显子 5 和第一个外显子 6 变异体(外显子 6.1)之间。在每个外显子 6 变异体前面都有一个拷贝的第二种元件选择子序列(selector sequence)(图 13-20 中以外显子 6.21 为例)。每一个选择子序列都是不同的,但如图 13-21 所示,它们都能与停泊位点配对。每个选择子序列结合停泊位点的区域有重叠,因此不同的选择子序列和停泊位点结合是互不相容的,即一次只有一个选择子能结合停泊位点。能结合到停泊位点的选择子序列使其相应的外显子 6 靠近外显子 5,从而确保该外显子 6 变异体被选择。

除了使被选择的外显子 6 变异体靠近外显子 5 外,选择子序列和停泊位点的杂交结合也确保该外显子 6 变异体免于普遍抑制机制的作用,这种机制可以抑制其他可能的外显子 6 变异体的剪接。一种蛋白(Hrp36)作为一种普遍剪接抑制因子通过覆盖其他外显子从而抑制其留在成熟 mRNA 上。由 RNA 杂交产生的局部抑制解除可能是 RNA 二级结构的直接结果;或者通过在 RNA 上形成一个结构,此结构可被移除抑制因子的蛋白所识别;或者使被选择的外显子靠近外显子 5 内的激活因子从而克服受到的抑制。

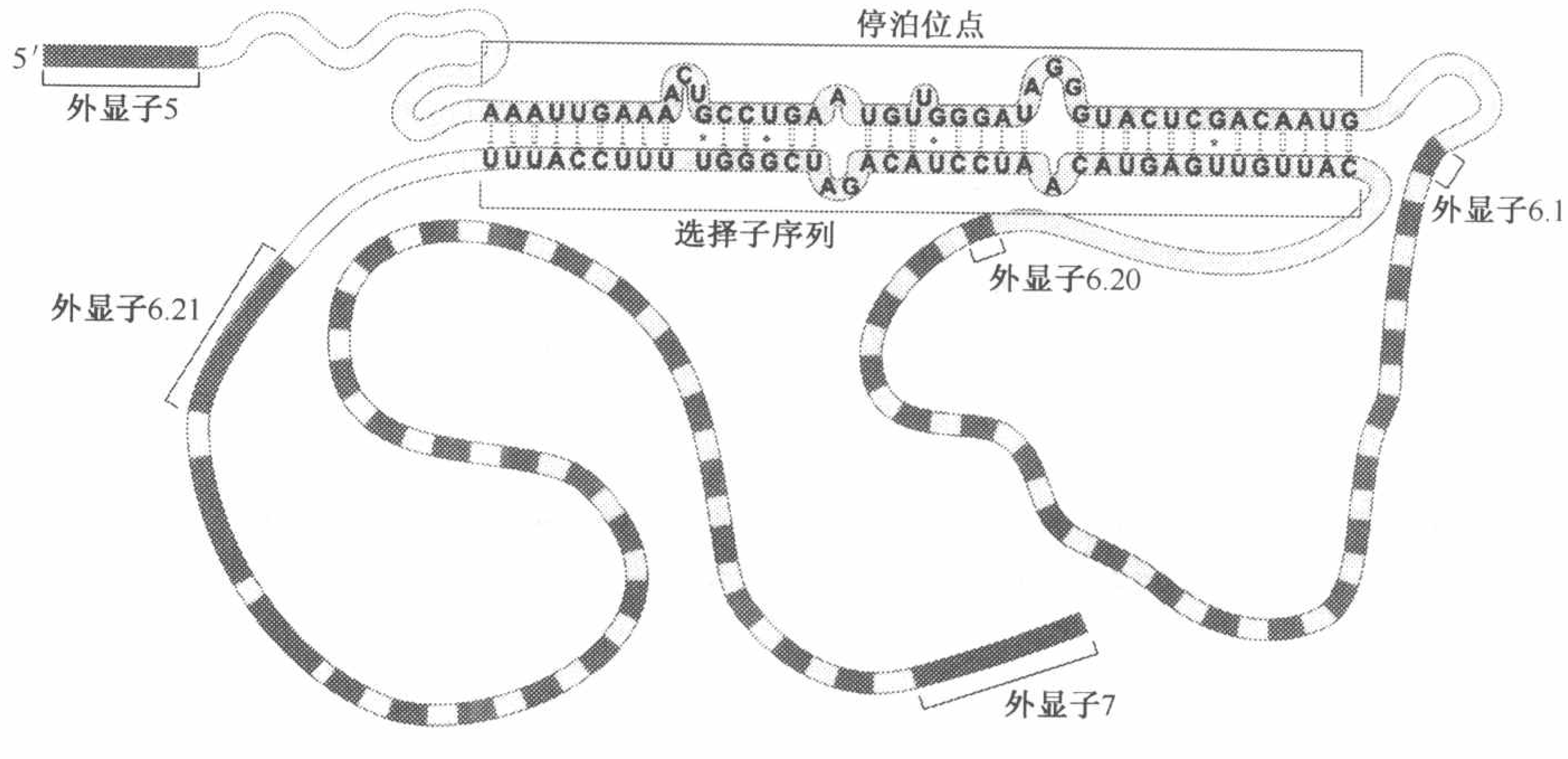


图 13-20 停泊位点:选择子序列。外显子 6.21 前的选择子序列与停泊位点的碱基配对。(Brenton Graveley 绘制、惠赠.)

框 13-3 描述了一个生物信息学分析的实例,正是在这个例子中,人们通过序列比对发现并鉴定了停泊位点和选择子序列。

可变剪接受激活因子和抑制因子调控

剪接调控蛋白结合到称为外显子/内含子剪接增强子(ESE 或 ISE,exonic/intronic splicing enhancer)或者外显子/内含子剪接减弱子(ESS 或 ISS,exonic /intronic splicing silencer,ISS)的特殊序列上。前者增强附近的剪接位点的剪接,后者正好相反。我们已经介绍过增强子及与其结合的 SR 蛋白(图 13-12)。这些序列和蛋白质对将剪接体导引到各个外显子区域非常重要,即使没有可变剪接也如此。在刚刚提到的组成型可

变剪接的例子中（图 13-16），也是 SR 蛋白保证了可变剪接的进行。该蛋白质家族数量庞大种类多样，在调控可变剪接方面也有特殊作用，即在不同的条件下引导剪接体到不同的剪接位点发挥作用。所以在发育的某个阶段，或者在某种类型的细胞中，一种特定的 SR 蛋白的存在与否或者活性高低，就可以决定某一特定的剪接位点是否得到利用。图 13-22 是假设的关于剪接调控的例子，激活因子结合到剪接增强子上，抑制因子结合到剪接减弱子上。

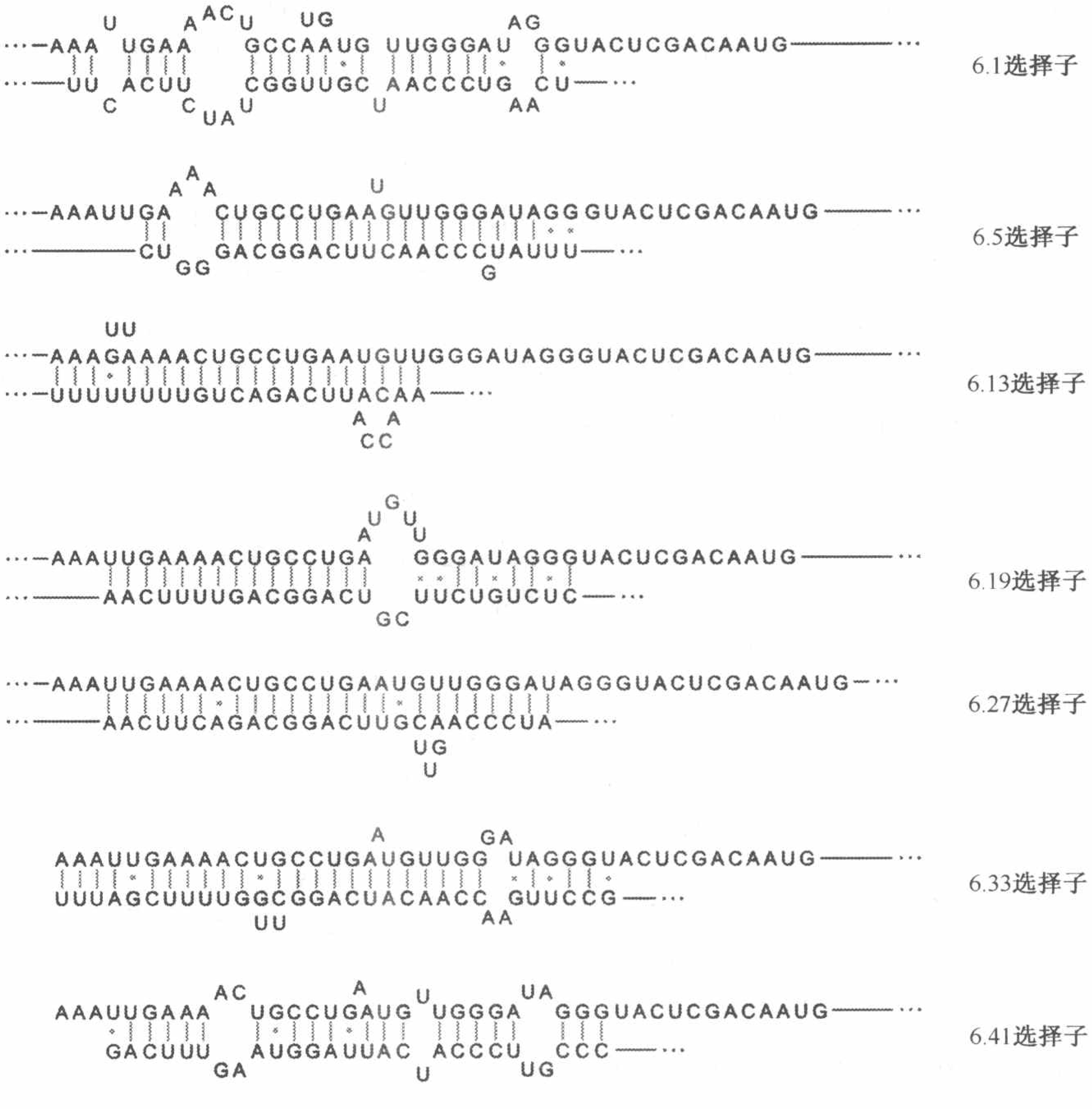


图 13-21　六个外显子 6 变异体的选择子序列，每个都与外显子 6 的停泊序列结合。每个选择子序列通过碱基配对与停泊序列的细微不同区域结合，但是这种与停泊位点的结合是互不相容的。（来源：Courtesy of Brenton Graveley.）

SR 蛋白利用某个结构域结合 RNA，如熟知的 RNA 识别基序（RNA-recognition motif，RRM）（图 5-23 及图 5-24）。每个 SR 蛋白都含有另一个结构域，富含精氨酸和丝氨酸，称为 RS 结构域（RS domain）。该结构域位于肽链的 C 端，介导 SR 蛋白与剪接体蛋白的相互作用，把剪接体募集到附近的剪接位点。

有一类激活因子能在特殊的组织中增强某种特定的剪接方式，果蝇的 Half-pint 蛋白即是。这个激活因子调控果蝇卵巢中一组 pre-mRNA 的可变剪接。它通过结合到那些 pre-mRNA 特定外显子的 3′剪接位点附近并募集 U2AF 因子来进行调控。

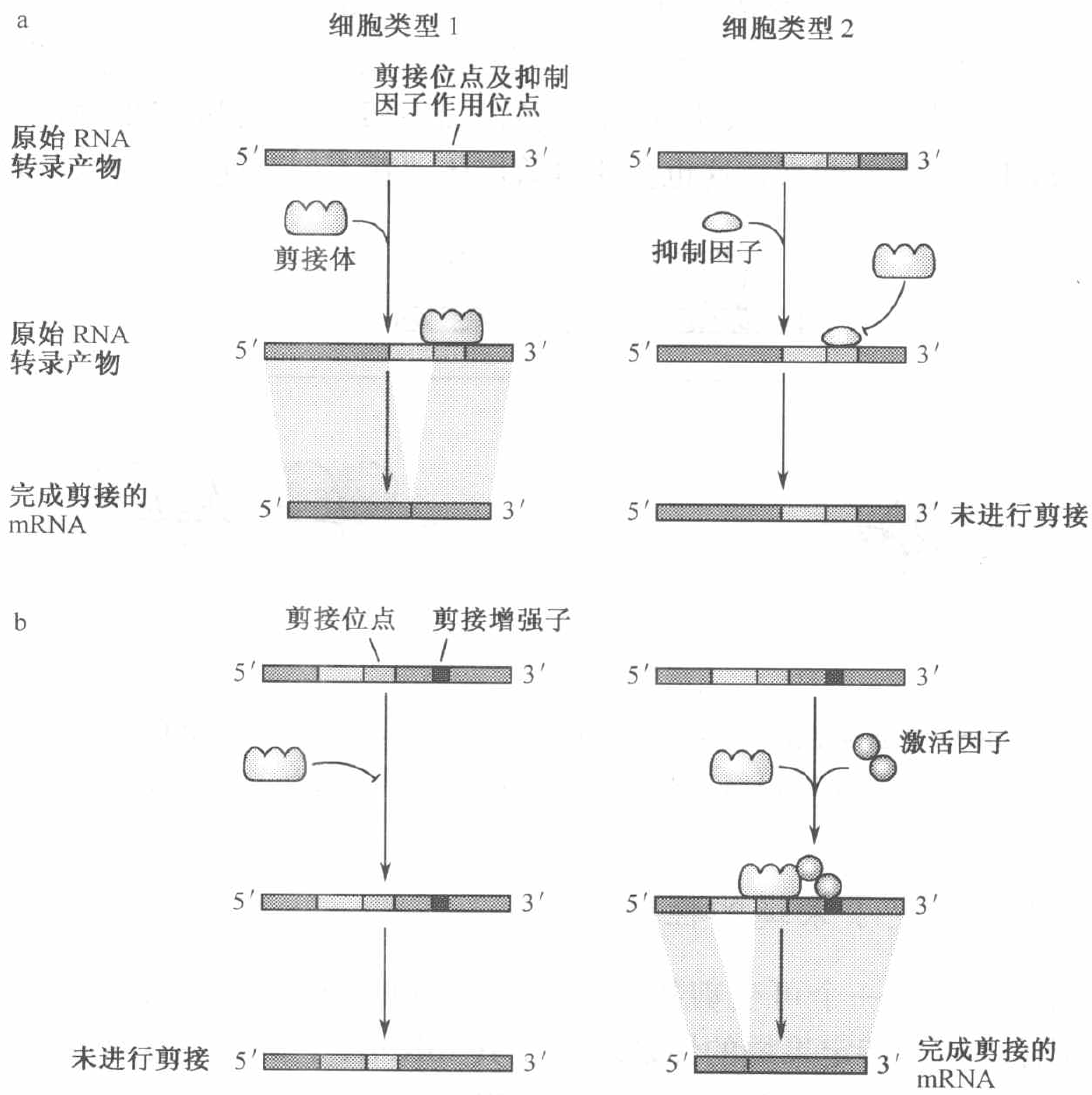


图 13-22 可变剪接的调控。某些可变剪接的外显子总是在成熟 mRNA 出现，除非受到某种抑制蛋白的阻止（见 a）；另外一些则相反，只有某种激活因子发挥作用，才能包含在成熟的mRNA中（见 b）。上面两种方式都可以用来进行剪接调控，使得某一特定的外显子出现在一种类型细胞的成熟 mRNA 中，而不会出现在另一种类型细胞的成熟 mRNA 中。

多数沉默子由不均一核糖核蛋白（heterogeneous nuclear ribonucleoprotein，hnRNP）家族的成员识别。这些蛋白质能结合 RNA 但没有 RS 结构域，所以无法募集剪接体。相反它们可以阻断特定的剪接位点，使之丧失作用我们在前面的 Dscam 例子中看到过类似的功能，Hrp36 抑制了外显子 6 的变异体留在 mRNA 中。另一个例子是，hnRNPA1，它可以识别并结合位于 HIV tat pre-mRNA 一个外显子内的沉默子并抑制该外显子最终出现在成熟的mRNA中。通过结合到该位点上，抑制因子阻断了激活因子 SC35（一种 SR 蛋白）对附近增强子的结合。这种阻断不是直接的（两个结合位点并不重叠），而是促进了其他的 hnRNPA1 协同结合到邻近的序列上，并整个覆盖了增强子位点。即便如此，另一个 SR 蛋白（SF2/ASF）仍然能够克服这种抑制作用，因为 SF2/ASF 对增强子的亲和力比 SC35 要大，所以能取代已经结合的抑制因子。在后面第 16、17 章关于转录调控的例子中，我们还会看到类似的协同和竞争结合情况。

哺乳动物的另一个剪接抑制因子是 hnRNPI 蛋白。有时 hnRNPI 直接结合到多聚嘧啶区（所以 hnRNPI 也叫多聚嘧啶区结合蛋白），阻断剪接体的结合。有时它结合到某个

外显子两个外侧的序列上，使该外显子不能进入成熟的 mRNA。单个外显子的剔除有两种可能的原因：①该外显子两端的 hnRNPI 通过相互作用，使外显子突起成为一个环，当剪接体经过时就把它漏掉了；②该外显子两端的 hnRNPI 与其他 hnRNPI 协同结合，将外显子所处的一段 RNA 覆盖起来，这也会使得剪接体“看不到”这个外显子（图 13-23）。

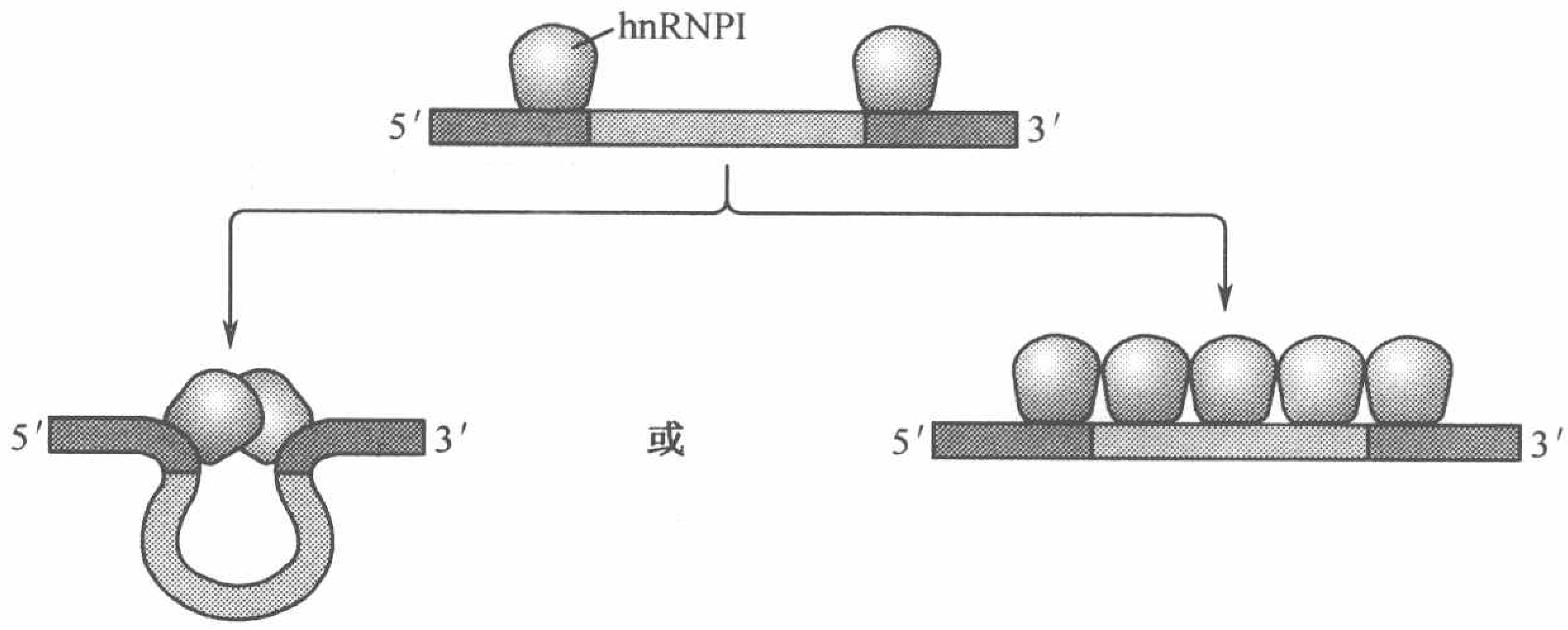


图 13-23 hnRNPI 对剪接的抑制作用。这里演示了两种模型，一种是 hnRNPI 蛋白把整个外显子包裹起来；另一种是 hnRNPI 结合到外显子的两端，使之突起成为一个环而隐蔽起来。

可变剪接的调控决定了果蝇的性别

现在，我们来探讨一个可变剪接调控非常精细的例子——果蝇 *double-sex*（*dsx*）基因的剪接调控。一只特定果蝇的性别取决于其 mRNA 上两个可变剪接产物中哪一个被制造出来。通过抑制因子和激活因子对一连串剪接事件（splicing cascade）的调控是决定果蝇性别的关键。

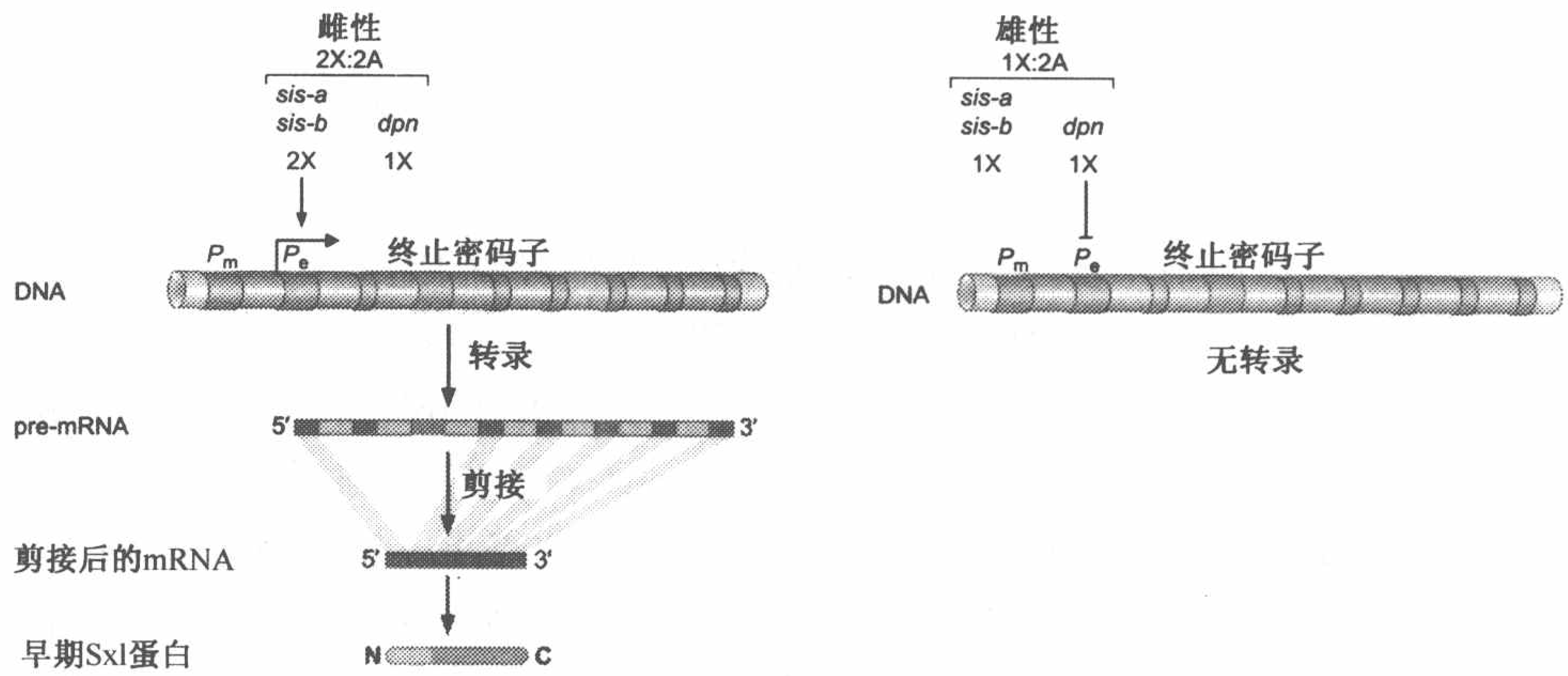


图 13-24 在雄性和雌性果蝇中 *Sxl* 基因的早期转录调控。*Sis A* 和 *Sis B* 基因位于 *X* 染色体上，编码控制 *Sxl* 基因表达的转录激活因子。Dpn 是 *Sxl* 基因的一个转录抑制因子，由染色体 2 上的一个基因编码。尽管雄性和雌性表达等量的常染色体编码的 Dpn，但是雌性产生 2 倍于雄性的转录激活因子（因为雌性有 2 条 *X* 染色体，而雄性仅有 1 条）。激活因子和抑制因子的比例不同保证了 *Sxl* 基因在雌性中表达，而在雄性中不表达。然后 *Sxl* 蛋白自调控 *Sxl* 基因的表达，如正文及下图中所述。（据 Estes P. A. et al. 1995. *Mol. Cell. Boil.* 15：904-917. 修改）

X 染色体和常染色体的比例决定果蝇的性别。X 染色体和常染色体的比例为 1（两条 X 染色体和两套常染色体）时，果蝇为雌性，比例为 0.5 时为雄性。这个比例最初用 *SisA* 和 *SisB* 两个转录激活因子的转录调控水平来检测（第 16 章和第 17 章将详细讨论转录调控机制）。编码这两个转录激活因子的基因都在 X 染色体上，所以在早期胚胎发育中，准雌性果蝇产生的 SisA 和 SisB 转录激活因子的量是准雄性的两倍（图 13-24）。

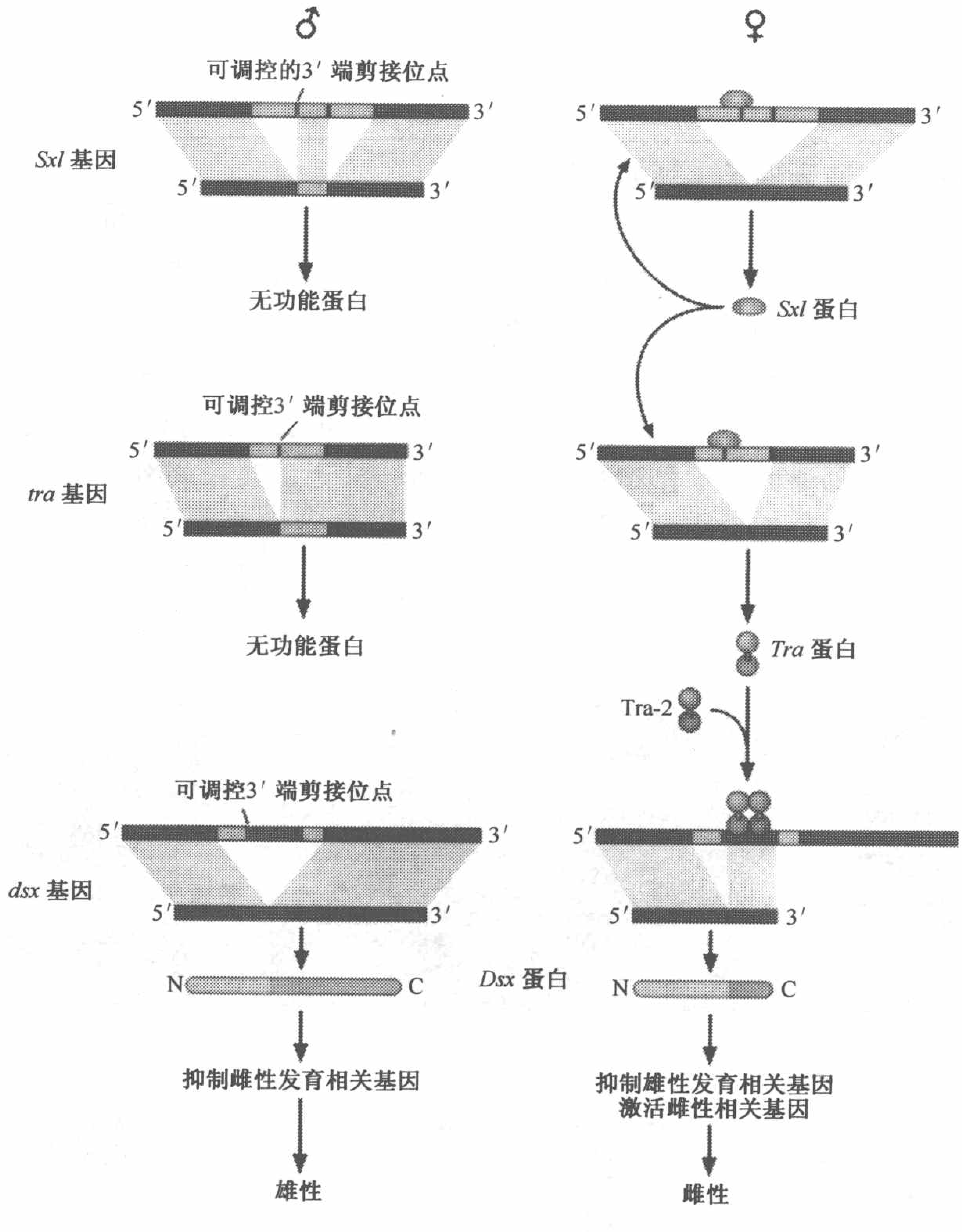


图 13-25 一连串的可变剪接决定了果蝇的性别。如正文中所详述，Sxl 蛋白在准雌性中表达（图右半部分），不在准雄性中表达（图左半部分）。Sxl 蛋白表达靠其自身调节其 mRNA 的剪接来维持。在雄性中没有这种调控，不产生有功能的 Sxl 蛋白。在雌性中 Sxl 蛋白还控制 *tra* 基因的剪接，产生有功能的 Tra 蛋白（雄性中无）。Tra 蛋白本身也是一个剪接调控因子，作用于 *dsx* 基因 pre-mRNA。*dsx* mRNA 在 Tra 蛋白的调控下剪接，产生的 Dsx 蛋白（雌性）其 C′端有 30 个氨基酸，与其在没有 Tra 蛋白（雄性）调控下产生的蛋白不同。在雌性中产生的 Dsx 蛋白激活雌性发育所需基因，抑制雄性发育所需基因。在雄性中产生的 Dsx 蛋白，其 C′端有 150 个氨基酸，抑制指导雌性发育的基因。Sxl 蛋白作为一个剪接抑制因子结合到 3′端剪接位点的多聚嘧啶区（图 13-3）。相反，Tra 蛋白作为一个剪接增强子结合 *dsx*RNA 某个外显子的增强子序列（图 13-12）。

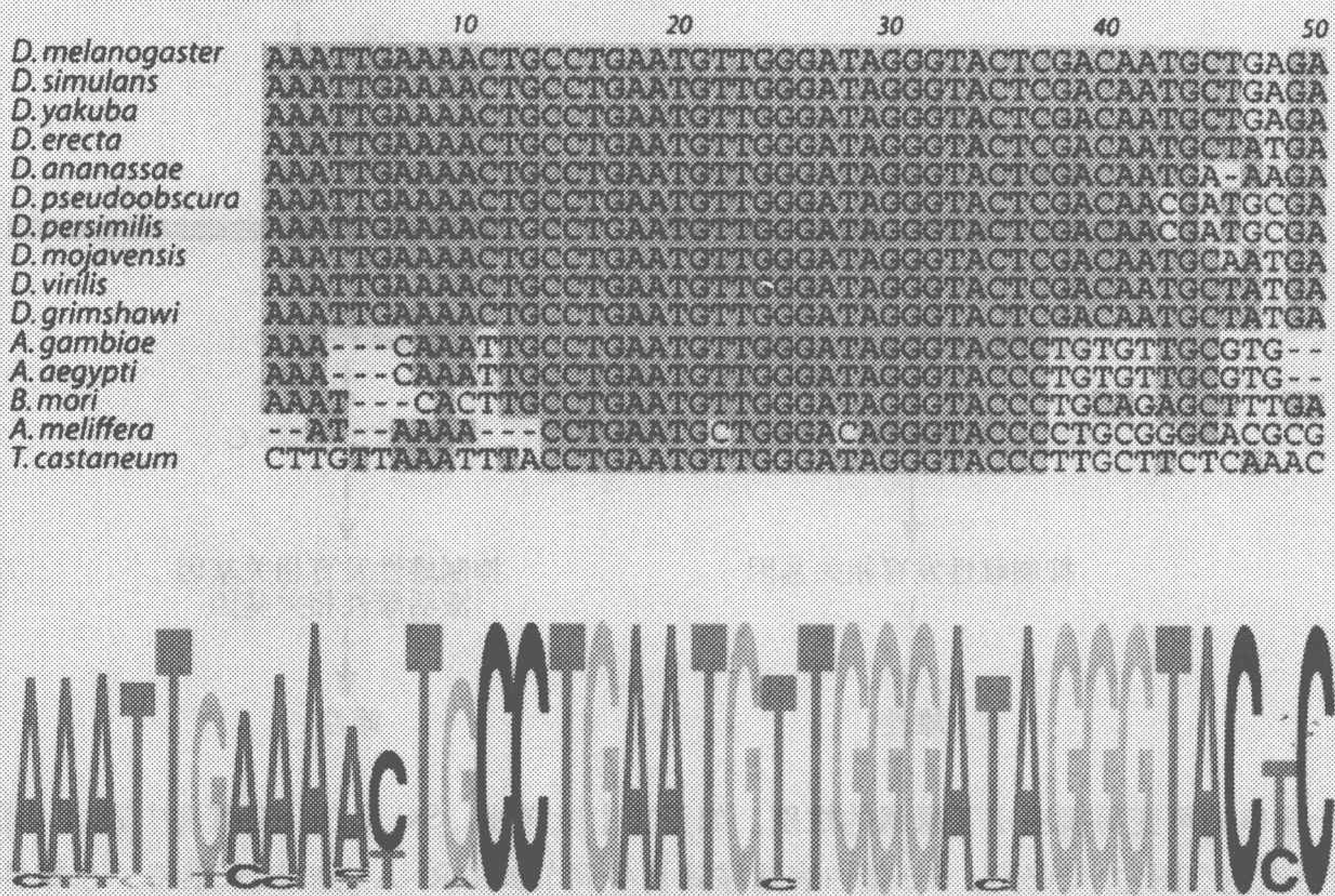
这些转录激活因子结合到 *sex-lethal*（*sxl*）基因调控序列上游的位点。另一个结合并控制 *Sxl* 基因转录的调控因子叫 Dpn（Deadpan），它是由常染色体（染色体 2）基因编码的。因此，转录激活因子和抑制因子在雌雄个体中的比例是不同的，这就造成了 *Sxl* 基因在雌性中被激活而在雄性中被抑制。

Sxl 基因的表达是从两个转录启动子开始的：P_e 和 P_m。P_e（promoter for establishment）受 Sis A 和 Sis B 调控，因此仅在雌性中表达。在发育后期，该启动子永久关闭。在雌性胚胎中，Sxl 的表达由 P_m（promoter for maintenance）来维持。

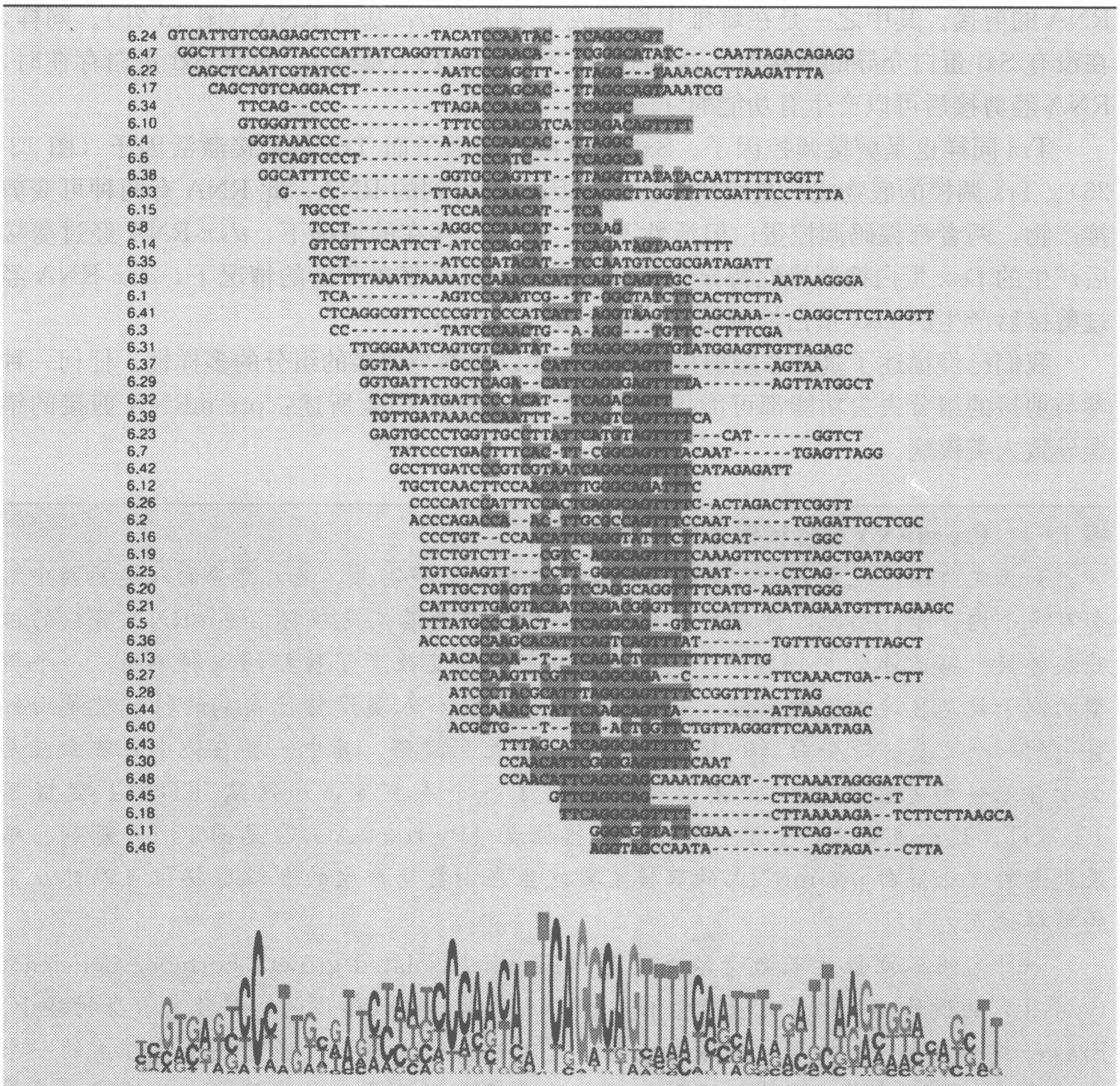
框 13-3 停泊位点和选择子序列的鉴定

停泊位点在黑腹果蝇中长 66 个核苷酸。在另外 10 种果蝇中，90%～100%的停泊位点序列是保守的。即使比对包含其他非果蝇昆虫（如蚊子、家蚕、蜜蜂），停泊位点中间的 24 个核苷酸仍然是高度保守的。事实上，它是整个 *Dscam* 基因（超过 60kb）中最保守的序列。停泊序列最初就是通过这种保守性鉴定出来的（框 13-3 图 1）。

选择子序列也是通过序列比对发现的，虽然其保守性不如停泊位点高。选择子序列是外显子 6 变异体上游内含子中相对保守的序列。对比黑腹果蝇外显子 6 变异体 48 个选择子序列，揭示了与停泊序列互补的一个 28 个核苷酸的共有序列（框 13-2 图 2）。当将每一个选择子序列与停泊位点比对时，发现每个选择子序列都可以与停泊位点配对，每个配对是唯一的，但相互重叠。实例见图 13-21。



框 13-3 图 1 十五种昆虫的停泊位点核苷酸序列比对。分析的昆虫包括 10 种果蝇；2 种蚊子，冈比亚疟蚊（*Anopheles gambiae*）和埃及伊蚊（*Aedes aegypti*）；家蚕（*Bombyx mori*）；蜜蜂（*Apis mellifera*）和赤拟谷盗（*Tribolium castaneum*）。每个停泊位点上最常见的核苷酸用阴影显示，共有序列用象形图在下部显示。每个字母的高度代表该核苷酸在此位置上出现的频率。（来源：Modified，with permission，from Graveley B. R. 2005. *Cell* 123：65-73，Fig. 2. © Elsevier.）



框 13-3 图 2 黑腹果蝇选择子的共有序列。图上部显示的是 48 个选择子及其两侧序列的比对。中间部分出现频率最高的核苷酸用亮标显示。图下部显示的是利用比对产生的选择子序列的共有序列。（来源：Reprinted，with permission，from Graveley B. R. 2005. *Cell* 123：65-73，Fig. 4. © Elsevier.）

无论是雌性还是雄性，P_m 起始的转录都是组成型的，但是由此产生的 RNA 比 P_e 起始转录得到的 RNA 多一个处显子。如果此外显子留在成熟的 mRNA 中，则无法产生有活性的蛋白，在雄性中即如此。但是在雌性中，通过剪接除去此外显子则可以持续地产生具有功能的 Sxl 蛋白。

如图 13-25 所示，正是在雌性中而不是在雄性中存在的（早期从 P_e 转录的）Sxl 蛋白介导由 P_m 转录的 mRNA 的剪接，确保有抑制活性的外显子被剪切掉。Sxl 蛋白作为一个剪接抑制因子执行上述功能。

因此功能性的 Sxl 蛋白可以持续地在雌性中产生。Sxl 蛋白调控其他 RNA 及自身

RNA 的剪接。其中之一是在雌雄中均组成型表达的 *tra* 基因 RNA（图 13-25）。同样，在没有 Sxl 蛋白介导的剪接时，这个 RNA 不产生蛋白（雄性），而当 Sxl 蛋白存在时，RNA 经剪接后可以产生有功能的 Tra 蛋白（雌性）。

Tra 同样也是剪接调控因子。Sxl 是剪接抑制因子而 Tra 是剪接激活因子（图 13-25）。Tra 调控的底物之一是编码 Dsx 的基因转录得到的 RNA。此 RNA 有两种可变剪接产物，两者都编码调控蛋白但活性不同。在 Tra 存在的情况下，*dsx* RNA 经过剪接后产生的 Dsx 蛋白抑制雄性特异性基因的表达。在不存在 Tra 的情况下，*dsx* RNA 经过剪接后产生的 Dsx 蛋白抑制雌性特异性基因的表达。

我们已经描述了真核系统剪接的各种方式以及参与剪接的组分的多样性。任何一种参与剪接的组分失去功能都可能导致严重后果。如框 13-4 所述：pre-mRNA 剪接的错误导致人类疾病。

框 13-4 Pre-mRNA 剪接错误导致人类疾病

如正文所述，绝大部分人类基因都含有内含子。事实上，绝大部分的人类基因中都含有多个内含子。因此，许多导致人类疾病的点突变其实是破坏 pre-mRNA 剪接的核苷酸替换。据估计至少 15%的导致人类疾病的点突变改变了剪接的识别序列。一个典型的例子就是β-地中海盆血症（β-thalassemia）。这种人类遗传性疾病的特征就是血红蛋白的一个亚基β-球蛋白（β-globin）的产生发生了障碍。其中一类β-地中海贫血症是由于β-球蛋白基因的第一个内含子中发生了一个点突变，序列从 TTGGT 变成了 TTAGT。这个点突变产生了一个 3′端剪接位点（Py tract AG/G 见图 13-3）序列。结果患者的β-球蛋白 pre-mRNA 的剪接主要发生在由突变产生的 3′端剪接位点而非正常的剪接位点。

一种名叫孤立性生长激素缺乏Ⅱ型（Familial isolated growth hormone deficiency type Ⅱ）的遗传病是由于生长激素 pre-mRNA 的剪接错误造成的，导致患者身材矮小。Frasier 综合征是一种泌尿生殖系统疾病，是对肾脏和生殖腺的发育起重要作用的一个基因的 pre-mRNA 剪接发生错误引起的。另外两个例子是一种痴呆症，是编码细胞骨架蛋白和一种囊肿性纤维化的 mRNA 剪接发生错误引起的。

还有其他疾病是由破坏剪接机器本身的突变造成的。例如色素性视网膜炎（rentinitis pigmentosa RP），该病的特征是视网膜进行性恶化直至失明。多个基因发生突变导致色素性视网膜炎，其中大部分基因具有视网膜特定的功能。但也有一些突变发生在编码剪接体组分的基因上。由于患者同时具有一份正常的基因拷贝和一份发生突变的基因拷贝，故其体内可以产生剪接蛋白，只是低于正常水平。

为什么一个剪接体组分低于正常水平所产生的影响只出现在一个特定的组织——视网膜呢？一种可能的解释基于这样一个事实，即视网膜上的感光色素视紫红质（the photopigment rhodopsin）需要高水平的周转。因此，剪接机器必须满足产生大量视蛋白（视紫红质的蛋白成分）的需求，以替代被降解的视蛋白。因此，视网膜可能对剪接的稍微改变比其他组织更敏感，因为其他组织不需要产生大量的特定蛋白。

脊髓肌肉萎缩症（spinal muscular atrophy，SMA）是一种在儿童中最常见的遗传性致死疾病。其特征是脊髓神经元的进行性退化，该病是由于运动神经元存活（Survival motor neuron，SMN）基因发生突变引起的。SMN 是剪接机器中普遍存在的组分，其准确的功能不是很清楚。与色素性视网膜炎的例子一样，我们不知道为什么剪接缺损的影响主要呈现在运动神经元。

上面介绍的例子都是遗传性疾病。但是这种剪接损坏导致疾病的突变同样发生在体细胞中。例如细胞周期调控因子 p73 基因的突变。p73 蛋白因其 mRNA 的可变剪接存在多种形式。导致 p73 pre-mRNA 可变剪接错误的突变与一种名叫鳞状细胞癌（squamous cell carcinoma）的癌症相关。导致很多其他 pre-mRNA 的剪接损坏或者可变剪接错误的体细胞突变很可能也是癌症病因之一。

人们认为医学是生物学最伟大的老师。当然，正如我们前面所见，这句格言恰好适用于 pre-mRNA 剪接领域，人类遗传疾病的研究对于探索剪接相关的序列以及剪接机制提供了巨大的启示。事实上，pre-mRNA 剪接机器最基本的组分 snRNPs 的发现正是基于一种自体免疫狼疮的研究，这种疾病的患者能产生针对自身核蛋白颗粒的抗体。对人类遗传疾病的持续研究很可能会对 pre-mRNA 的成熟机制有更多启示。

外显子改组（Exon Shuffling）

外显子通过重组的方式完成改组，从而产生了编码新蛋白质的基因

如我们所看到的，所有真核生物都含有内含子，然而这种结构元素在细菌中却极为罕见——几乎不存在。这种情形有两种可能的解释。

第一种为内含子在先模型（intron early model），即内含子曾存在于所有生物，但在细菌中丢失了。假设细菌中确实原本就存在内含子，那么内含子为什么会在后来丢失呢？理由是这些“富含基因”的生物（第 7 章和第 11 章）为了应对环境的压力，精简了基因组结构以提高染色体复制和细胞分裂的速度。（真核生物中，单细胞且生长较快的酵母，其内含子数目比复杂的多细胞生物要少。）

另一种解释为，内含子在细菌中从来就不存在，而是进化过程中后来才出现的。按照这种所谓的内含子在后模型（intron late model），内含子可能通过类似转座子的机制（第 11 章），插入到原先没有内含子的基因之中。

不管孰是孰非，目前还不能明确解决这个问题，但现在提出第二个，也许是更有意思的问题：为什么内含子在真核生物中得以保存下来，尤其是在多细胞真核生物中那么广泛的形式存在？内含子存在的一个明显好处是，满足在剪接过程中除去它的需求，使可变剪接成为可能，从而可以由一个基因得到多种蛋白质产物。但从更宏观的角度来看，有利于这些生物的另一个好处是：把编码区分成几个外显子使得通过外显子改组的方式产生新基因成为可能。有三个观察结果有力地证明这个过程确实发生。

• 首先，某个特定基因的外显子和内含子的交界处经常与该基因编码蛋白质的结构域（第 5 章）的边界相一致，即每个外显子似乎经常编码蛋白质的一个独立的折叠单位（也经常与一个独立的功能相一致）。以图 13-26 所示的 DNA 结合蛋白为例。与多数 DNA 结合蛋白一样，这个蛋白质有两个结构域：DNA 识别区和二聚化区。如图所示，这两个结构域（D1 和 D2）分别由基因中不同的外显子（E1 和 E2）编码。

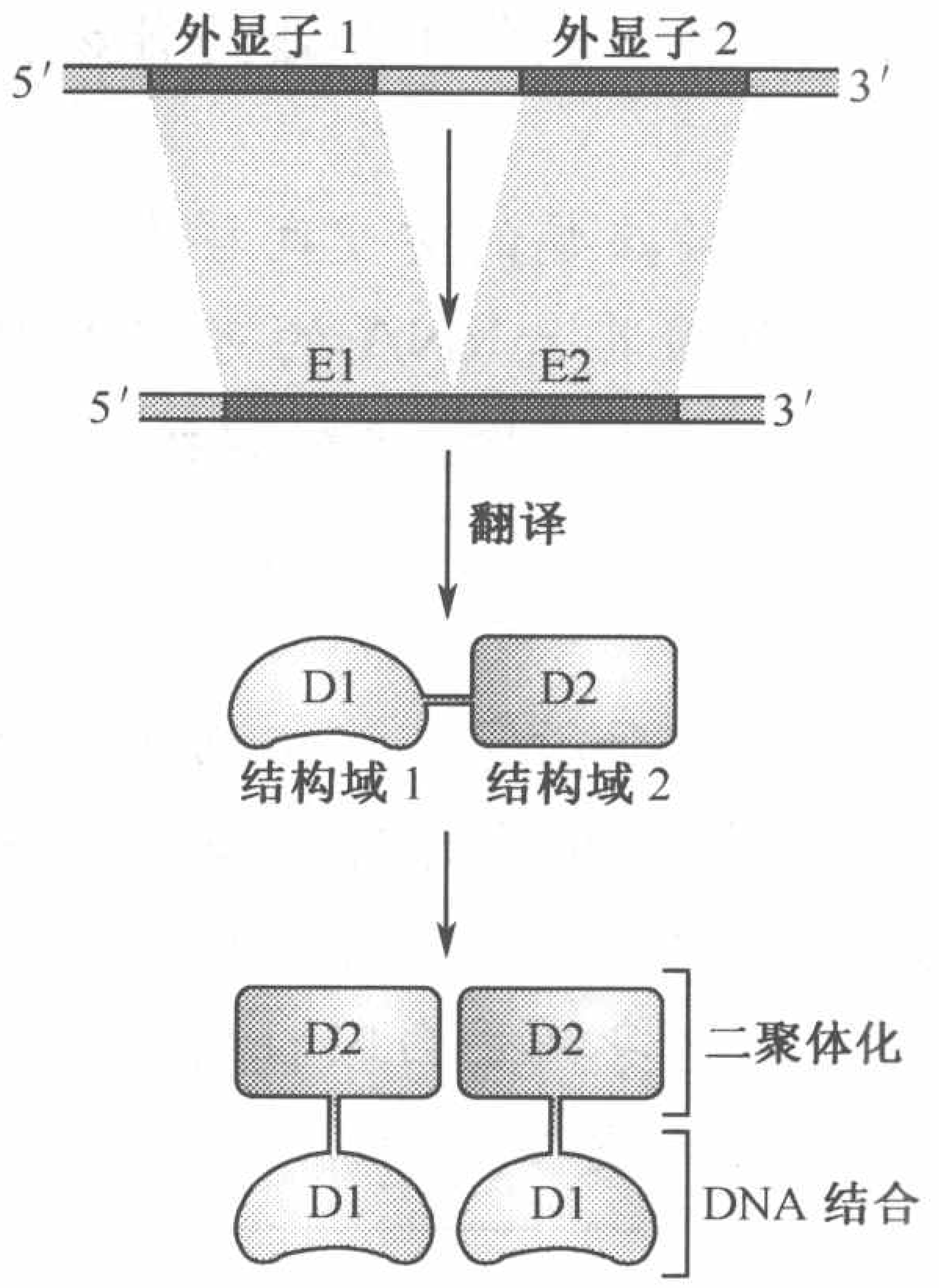


图 13-26 外显子编码蛋白质结构域。在本例中蛋白质的 DNA 结合区由一个外显子编码，同一蛋白质的二聚体化结构域由另一个外显子编码。蛋白质结构域独立进行立体折叠，不依赖于它所存在的蛋白质的其余部分，而且通常执行单一的功能（如我们第 5 章所讨论的）。这样，外显子经常可以在蛋白质之间活跃地交换。

• 其次，许多基因及其所编码的蛋白质明显是进化过程中部分地通过外显子复制和变异而产生。由多个重复单位构成的蛋白质（如免疫球蛋白）很可能这样起源（第 11 章图 11-35）。每个外显子之间都存在内含子，使得重复更有可能。

• 第三，有时在其他方面并不相关的基因中会发现相关的外显子，即有证据显示有些外显子确实在编码不同蛋白质的基因中被重复使用。以低密度脂蛋白（low-density lipoprotein，LDL）受体基因为例（图 13-22）。该基因的部分外显子在进化上明显与编码表皮生长因子（epidermal growth factor，EGF）前体基因的外显子有关。同时它又有一些外显子明显与补体成分 C9 基因的外显子相关（图 13-27）。有关外显子增加的更多例子在基因组全序列（如人类基因组）中非常明显。如图 13-28 所示，外显子以改组方式组合成不同基因，编码出的蛋白质由高度相关的结构域以各种方式组合而成，这方面的例子不胜枚举。

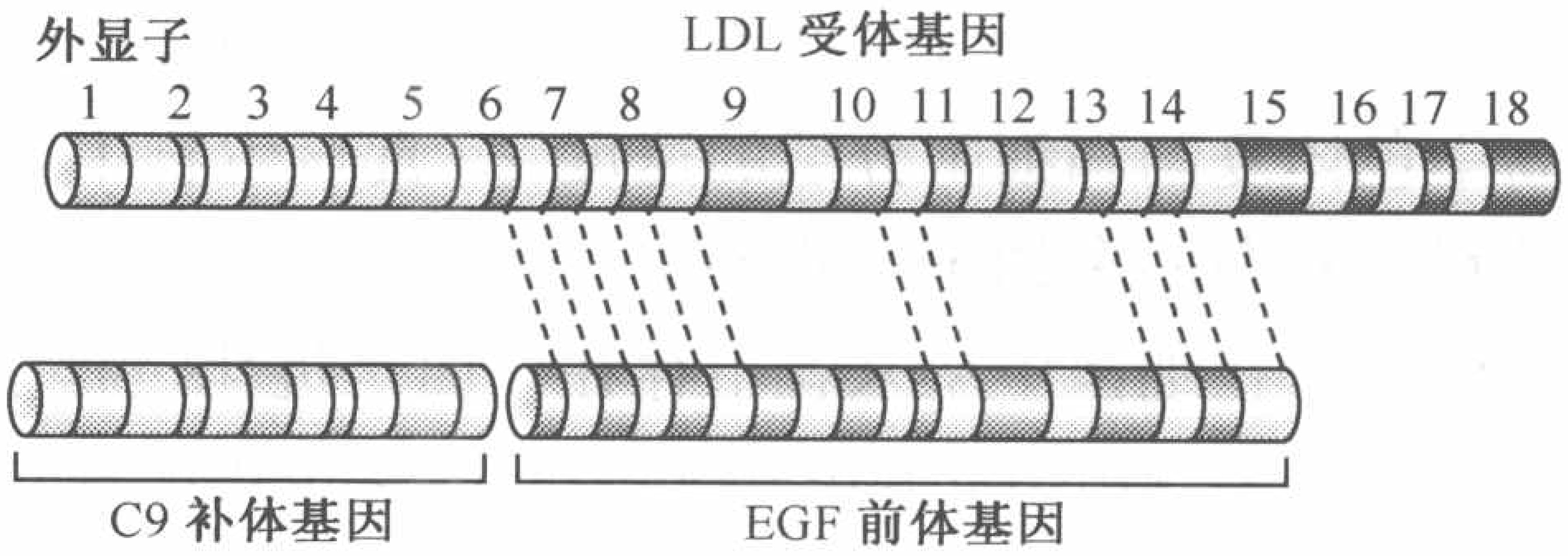


图 13-27 由其他基因的部件构成的基因。LDL 受体（血浆低密度脂蛋白受体）基因包含与 C9 补体基因密切相关的一串 6 个外显子，以及与 EGF（表皮生长因子）前体基因密切相关的 8 个外显子。因此，LDL 受体基因是由其他基因的外显子经过改组得到的；虽然这里没有列出，但同一批外显子也出现于其他基因中。在多数情况下，内含子在 EGF 前体基因和 LDL 受体基因中并非严格地处在同一对应位置上。在本图中，对应位置相同的内含子以虚线表示。

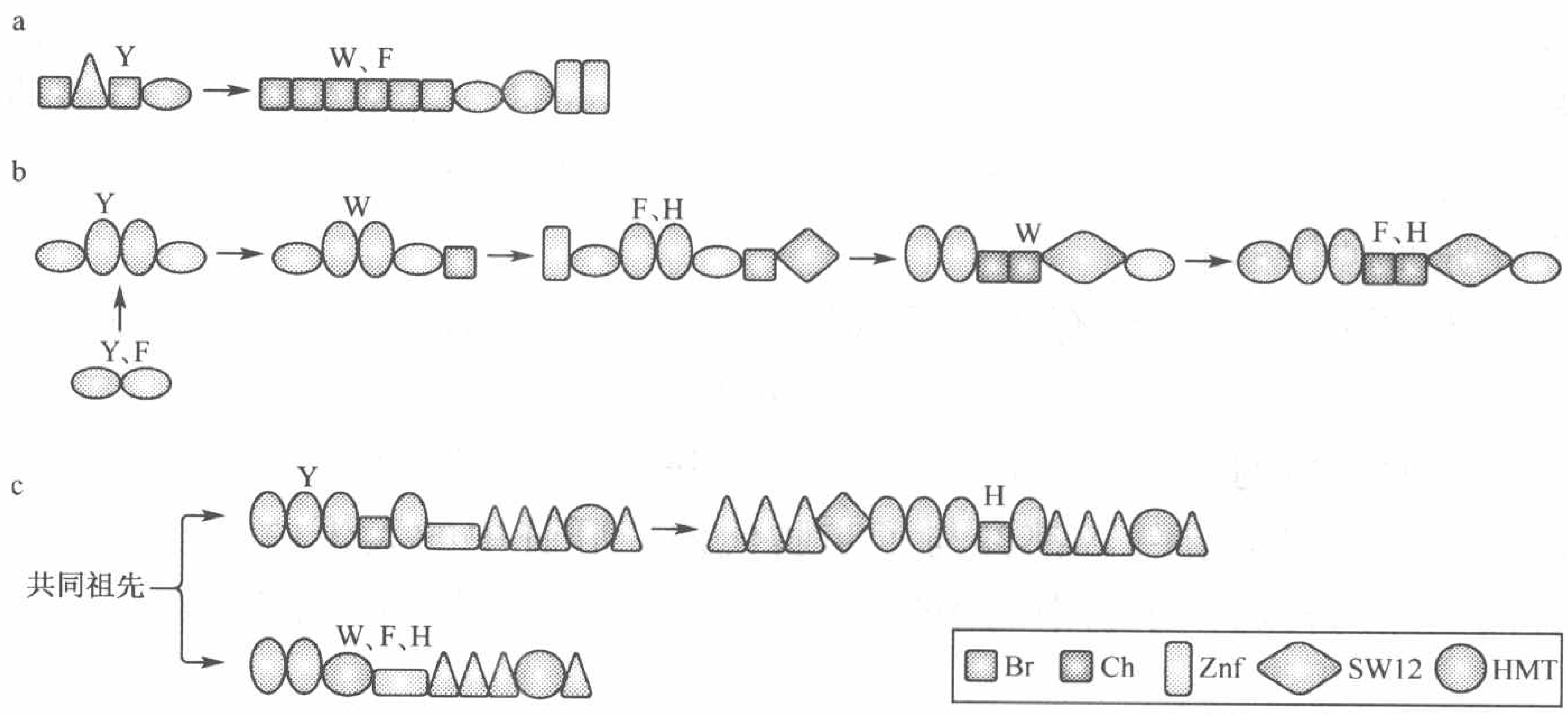


图 13-28　在某蛋白质家族的进化过程中结构域的累积、丢失和改组。本图示意了不同的相关蛋白质通过特定结构域的得失而进化产生的假设途径。给出了三个例子，都是关于染色质修饰酶（第 7 章），分别来自酵母（Y）、蠕虫（W）、昆虫（F）和人类（H）。每种蛋白质都以不同颜色和形状的一系列结构域来显示，每个蛋白质的上方标明了存在该形式结构域组合的物种。有些组合不止存在于一种生物体内，而且一种生物有时会存在类似结构域的多种相关组合。少数结构域是完全一样的，其功能在第 7 章或第 17 章曾讨论过：bromodomain（Br）、染色质结构域（Ch）、组蛋白甲基转移酶结构域（HMT）、与染色质重组酶有关的 ATP 酶活性（SW12）和锌指结构域（Znf）。（来源：Adapted，with permission，from Lander et al. 2001. Nature 409：p. 906，Fig 42. © Macmillan）

如我们已经看到的，外显子一般很短（约 150nt），而内含子长度不一，实际上可以很长（可达几百 kb）。对高等真核生物的多数基因而言，这样的长度比例可以保证重组更多地出现于内含子中而不是在外显子区，因而外显子更可能进行改组而不是被中断。剪接的机制——使用 5′和 3′剪接位点，保证了几乎所有重组基因都可以得到表达，因为不同基因的剪接位点很大程度上可以互换。另外，可变剪接可以在不改变原基因产物的基础上尝试新的外显子，也就是说，一开始新旧产物都可以产生。

RNA 编辑

RNA 编辑是改变 mRNA 序列的另一种方法

与 RNA 剪接一样，RNA 编辑（RNA editing）可以在 RNA 转录之后改变其序列。因而翻译出来的蛋白质与根据基因序列所推导出来的不同，但比剪接更富戏剧性——并非将上段 mRNA 进行重排，在 RNA 编辑时，个别碱基插入、删减、或改变。也就是说，RNA 中的编码信息发生改变。介导 RNA 编辑的机制有两种：位点特异性脱氨基作用和引导 RNA 指导的尿嘧啶插入或删除。我们依次进行讨论。

位点特异性脱氨基作用（deamination）的一种形式是，mRNA 中某个特异选择的胞嘧啶经过脱氨基变成了尿嘧啶。通常，对于一种给定的 mRNA，RNA 编辑只在特定的组织或细胞发生，而且受到调控。图 13-29 是哺乳动物载脂蛋白 B（apolipoprotein-B）基因。

这个基因有几个外显子，其中一个外显子的特定 CAA 密码子是 RNA 编辑的靶标，该密码子的胞嘧啶被脱氨基。由胞苷脱氨酶（cytidine deaminase）完成的脱氨基反应使胞嘧啶变成了尿嘧啶（图 13-30）。在这个例子中，脱氨基反应具组织特异性：mRNA 在小肠细胞而不在肝脏细胞中编辑。

CCA 密码子在肝脏中没有进行 RNA 编辑，翻译成谷氨酰胺；而在小肠中经过 RNA 编辑，变成了终止密码子 UAA。结果就是，全长的蛋白质（大约 4500 个氨基酸）在肝脏产生，而截短的多肽（只有约 2100 个氨基酸）在小肠里制造（图 13-29）。

这两种形式的载脂蛋白 B 都参与了脂肪代谢。肝脏中的大蛋白参与内源合成的胆固醇和甘油三酯的转运，小肠中的小蛋白参与运输饮食中的脂肪到各种组织中。

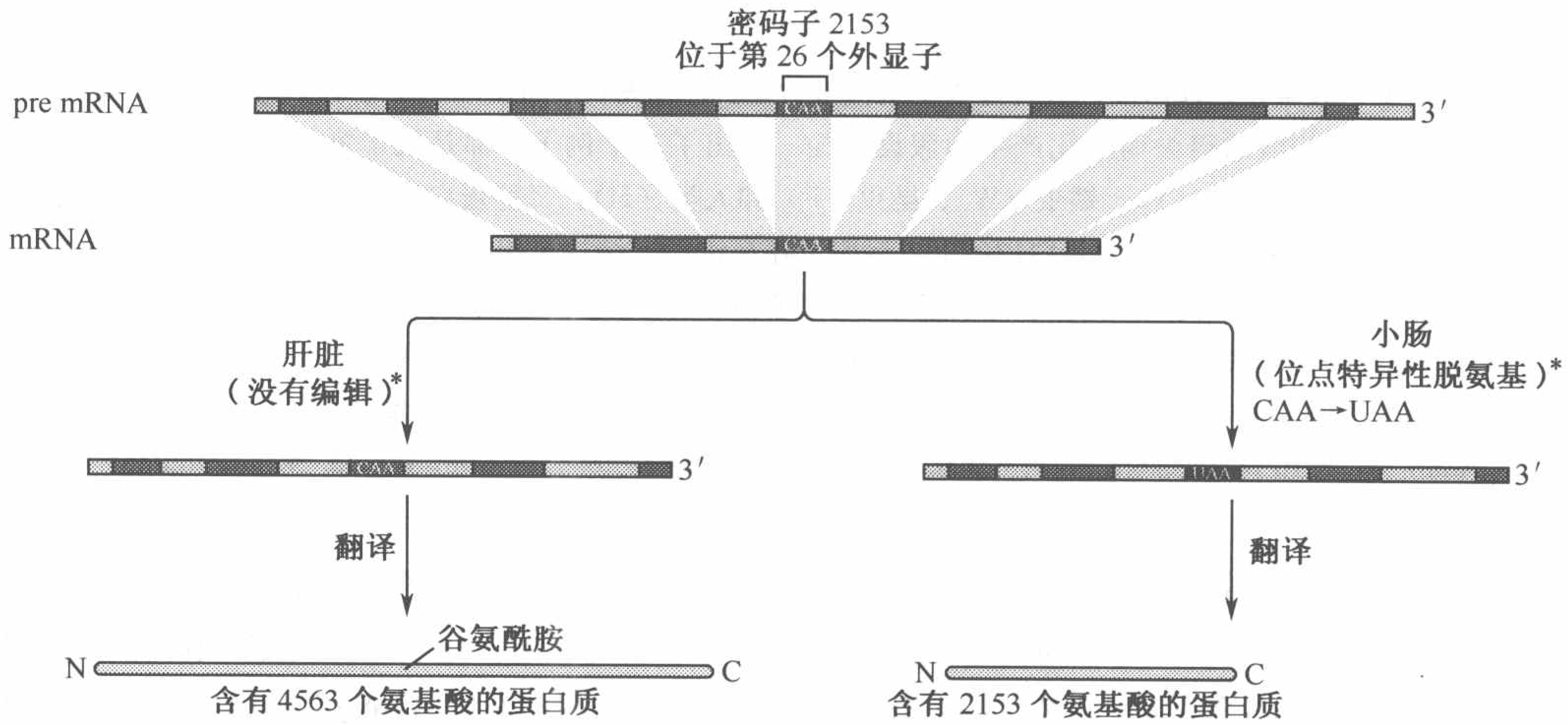


图 13-29 通过脱氨基方式进行 RNA 编辑。人载脂蛋白 RNA 以组织特异性的方式进行编辑，一个特定的胞嘧啶经过脱氨基变成了尿嘧啶。这个过程发生在小肠细胞的 RNA 中，而肝脏细胞中则没有。结果如正文所述，在小肠的 mRNA 中出现了一个终止密码子，所以翻译出的蛋白质分子比肝脏中的小。本图未按比例绘制：被编辑的是第 26 个外显子，填写在整个外显子中的密码子（CAA）实际只占该外显子的很小一部分。* 原文有转运到（transport to）一词，不妥。——译者注

其他酶促脱氨的 RNA 编辑例子包括腺嘌呤的脱氨基作用。这个反应由作用于 RNA 的腺苷脱氨酶（adenosine deaminase acting on RNA，ADAR，人类有三种）催化，产生次黄嘌呤。由于次黄嘌呤可以与胞嘧啶配对，所以这种编辑形式很容易改变 mRNA 编码的蛋白质序列。哺乳动物大脑中表达的一种离子通道蛋白是通过这种方式编辑的。对其mRNA的一个单一的编辑导致这种离子通道蛋白的一个氨基酸改变，从而改变了该通道 Ca^{2+} 的通透性。如果不发生这种编辑，大脑的发育就会严重受损。

酶促脱氨这类 RNA 编辑形式似乎很少见，但是很重要。在果蝇中，估计可能最多有 20 个胞嘧啶是脱氨基的靶点，但都存在于编码与神经递质产生或者活性相关的基因中。

脱氨酶作用的特异性还不完全清楚：脱氨酶的活性位点可以作用于任何胞嘧啶。通常脱氨酶以复合体形式起作用，复合体的其他组分可能影响脱氨酶作用的特异性。此

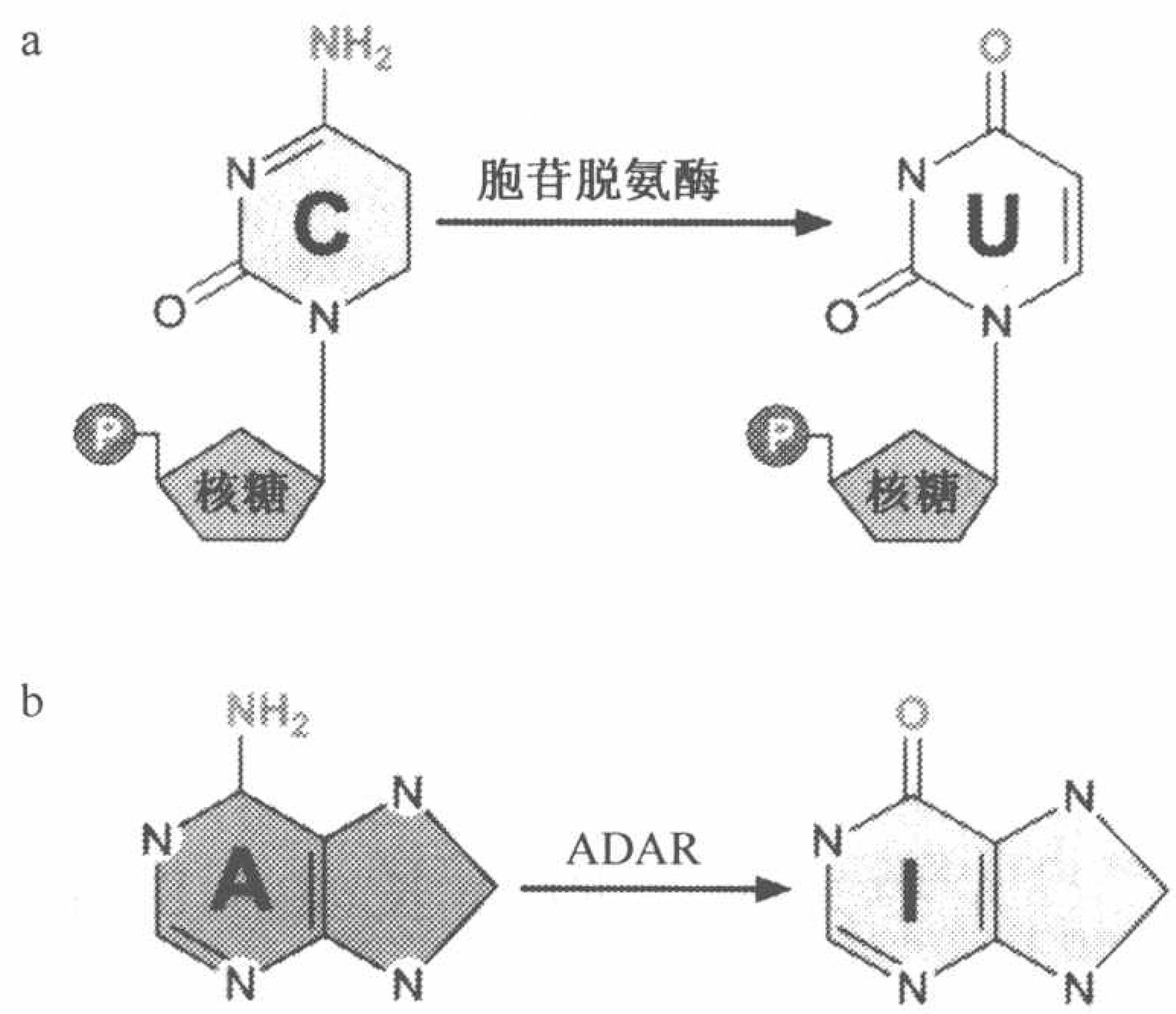


图 13-30 胞嘧啶和腺嘌呤脱氨基产生尿嘧啶和次黄嘌呤。a. 核苷酸环上的氨基基团被胞啶脱氨酶移走。b. 在腺嘌呤脱氨中，同一基团由 ADAR 移走以产生次黄嘌呤。

外，在上述的作用于载脂蛋白 B 的胞嘧啶脱氨酶例子中，脱氨酶上存在一个 RNA 结合区域，它通过识别 RNA 上一段特异序列或者特定的二级结构的方式帮助识别需要脱氨的特异位点。

在细胞抵抗 HIV 感染中脱氨酶的另外一个作用详见框 13-5：脱氨酶和 HIV。

引导 RNA 指导尿嘧啶的插入和删减

框 13-5 脱氨酶和 HIV

如正文所述，人载脂蛋白 B mRNA 的脱氨是由称为 APOBEC1（apolipoprotein-B editing enzyme，catalytic polypeptide-like 1）的酶进行的。APOBEC1 是 RNA 和 DNA 胞嘧啶脱氨酶家族中的一个成员。该家族的另一个成员 APOBEC3G（A3G）能有力地抑制一系列逆转录病毒包括 HIV 的感染。

逆转录病毒颗粒如 HIV 携带 RNA 基因组。一旦感染，其 RNA 通过逆转录产生一个 cDNA 拷贝（框 11-2），A3G 酶攻击的正是逆转录过程中产生的这个负链 cDNA。A3G 酶催化 cDNA 上胞嘧啶脱氨基而成为尿嘧啶，导致病毒无法修复的突变程度，甚至突变的链被 DNA 糖基化酶和 AP 核酸内切酶（apurinic-apyrimidinic endonuclease）所破坏（第 9 章）。

为了减少这个影响，野生型 HIV 产生 Vif（viral infectivity factor）蛋白，介导 A3G 酶的降解，从而将 A3G 酶排除在病毒颗粒之外，使病毒在下一轮复制中免受 A3G 酶的伤害。在体内，Vif 对于病毒在所有其生物相关的宿主细胞中的生长都是必须的。vif 缺陷型病毒可以在实验室使用的一些用于生长病毒的细胞系中生长，这些细胞被称作耐受细胞（permissive）。耐受细胞和非耐受细胞融合得到的异核体具有非耐受的特性。由此可见，非耐受细胞产生一种因子抑制病毒复制。这种因子就是脱氨酶 A3G，耐受细胞仅仅通过表达 A3G 蛋白即可转化为非耐受细胞。

在锥虫线粒体中发现了另一种截然不同的 RNA 编辑形式。在这种情况中，RNA 转录后，在 pre-mRNA 的特定位置插入多个尿嘧啶（有时也可能是删除几个尿嘧啶）。插入的数量可能非常大，最大时达到成熟 mRNA 核苷酸总数的一半左右。多个尿嘧啶的插入使得 mRNA 的密码子和阅读框发生了改变，从而完全改变了 mRNA 的“意思”。以锥虫 *coxII* 基因为例，在这个基因 mRNA 的特定位置上，3 个位点的比邻碱基间共插入 4 个尿嘧啶（其中一个位点 2 个，另外两个位点各 1 个）。这些插入改变了某些密码子，并使阅读框向前移了一个碱基，而这种移码突变正是得到正确阅读框所需要的（图 13-31a）。

这些额外的碱基是如何插入的呢？如图 13-31 所示，尿嘧啶是由引导 RNA（guide RNA，gRNA）插入到 mRNA 中去的。gRNA 长 40～80nt，是由不同于编码其靶 mRNA的基因所编码的。每种 gRNA 都可以分为三个区，第一区位于 5′端，称为“锚”区，负责指引 gRNA 到达它所要编辑的 mRNA 的目标区域；第二区用来精确定位在被编辑的序列中尿嘧啶将插入的位置；第三区位于 3′端，是一段多聚尿嘧啶序列。下面仔细看一下 gRNA 是如何指导 RNA 编辑工作的。

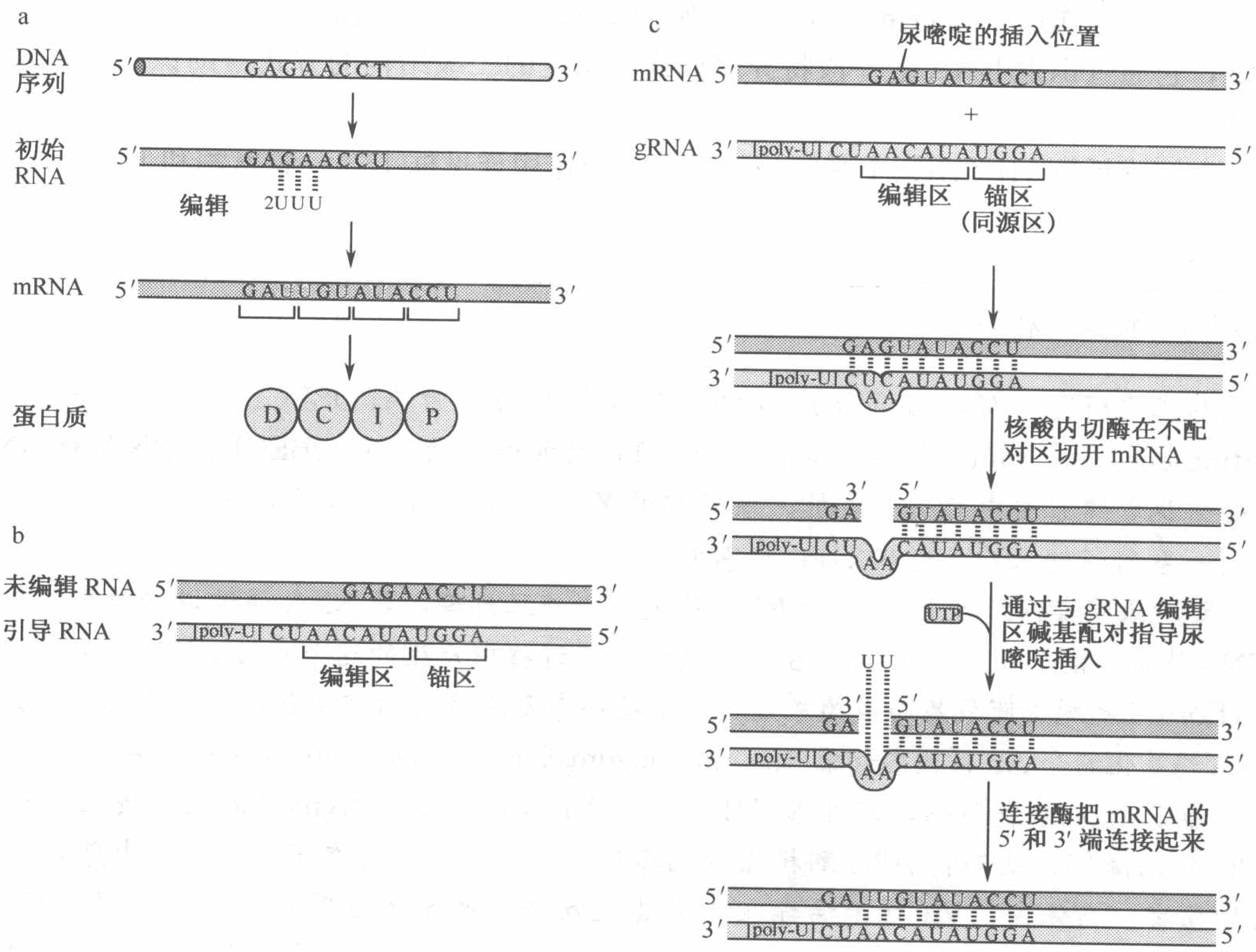


图 13-31 由 gRNA 介导的尿嘧啶插入法进行 RNA 编辑。锥虫 *coxII* 基因 RNA 的编辑。a. 示意 *coxII* 基因 pre-mRNA上 4 个尿嘧啶核苷酸插入的位置，在 mRNA 上形成正确可读框及编码信息；b. 标出了 gRNA 上决定尿嘧啶插入模式的序列，以及未编辑 mRNA 的序列；c. 演示了编辑反应的全过程。

gRNA 的“锚”区有一段序列，可以与 mRNA 上将要编辑区域的紧邻位置上（3′端）的一段序列形成碱基配对（图 13-31b）。随后编辑“指令”发出：gRNA 的一段序列与 mRNA 上的被编辑区互补配对，但有几个多余的腺嘌呤。这些腺嘌呤在 gRNA 上的位置与将要在 mRNA 上插入尿嘧啶的位置正好相反。gRNA 的 3′端是多聚尿嘧啶，该区域核苷酸的作用尚不清楚，但有人认为它可以把 gRNA 系于 mRNA 上编辑区上游（5′端）的富含嘌呤序列上。

如图 13-31c 所示，gRNA 与 mRNA 形成 RNA-RNA 双链，而在将要插入尿嘧啶的对应位置形成突出的单链环状结构。一种核酸内切酶识别并切开与环状结构相对的 mRNA。编辑涉及在 mRNA 上打开的缺口中加入尿嘧啶，这个过程由 3′端尿苷酰转移酶（TUT 酶）催化。

尿嘧啶插入后，某种 RNA 连接酶把两段 mRNA 连接起来，而 gRNA 的编辑区则沿 mRNA 从 3′→5′方向继续发挥作用。一条 gRNA 可以负责在不同位置插入多个尿嘧啶（图 13-31）。另外，有时会有几种不同的 gRNA 在同一条 mRNA 的不同区域发挥作用。

mRNA 转运

一旦完成加工，mRNA 会包装好并从细胞核输出到细胞质用于翻译

mRNA 一旦全部加工完毕（加帽、剪接和多聚腺苷酸化）就会运出细胞核进入细胞质（图 13-32），在此翻译得到蛋白质产物（第 14 章）。从细胞核进入细胞质的过程并非是一个被动的过程。事实上，它必须受到精细调控：完成加工的 mRNA 只占细胞核内全部 RNA 的很少一部分，许多其他 RNA 如果进入细胞质会对细胞产生致命的危害。这些其他 RNA 包括损伤或加工错误的 RNA，以及剪接下来的内含子等（由于内含子一般比外显子大得多，实际上也确实如此，所以与成熟 mRNA 相比，内含子占了细胞核 RNA 的绝大部分）。

RNA 选择性转运是如何完成的呢？正像我们在本章及前一章一直强调的那样，从 RNA 分子开始转录的那一刻起，它就与多种蛋白质结合在一起：最初是参与加帽反应的蛋白质，然后是剪接因子，最后是介导多聚腺苷酸化的蛋白质。其中一些蛋白质在加工过程的不同阶段被替换，而另外一些（如某些 SR 蛋白图 13-12）则不是；此外，还会有另外的蛋白质加入。结果是典型的成熟 mRNA 携带着一组蛋白质，正是它们决定了 mRNA 被转运的命运。而其他 RNA 不仅缺少转运所必需的特定标记性蛋白质组合，而且自身也携带另外一组蛋白质，这些蛋白质可以有效地阻止 RNA 转运。例如，剪接下来的内含子通常携带 hnRNP（剪接抑制酶，图 13-23），它们决定了这些 RNA 要留在细胞核内并被降解。

成熟 mRNA 除了携带残留的 SR 蛋白，还携带另外一组专门结合外显子接头的蛋白质（当然只见于完成剪接的 RNA）。mRNA 确实也含一些 hnRNP，但数量比典型地

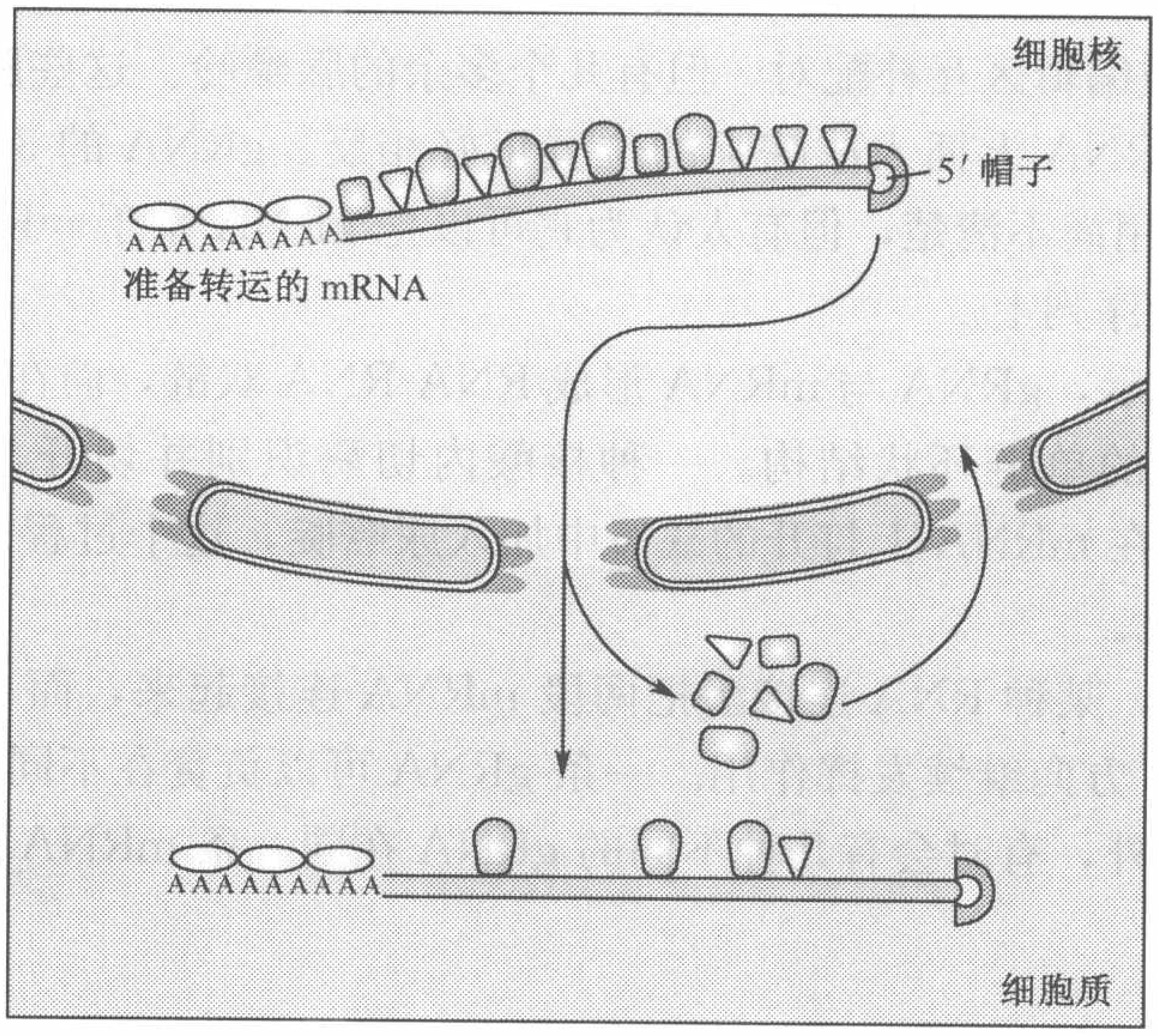


图 13-32 mRNA 转运出细胞核。RNA 运出细胞核是一个主动转运过程，只有某些合适的 RNA 才被挑选出来进行转运。为了被挑选转运，这类 RNA 必须结合一组正确的蛋白质。这些蛋白质会使之区别于其他必须留在细胞核内或必须降解的 RNA 种类。例如，识别外显子交界点的蛋白质表示 mRNA 已得到正确剪接，而结合内含子的蛋白质则确定了该种 RNA 应该留在核内。一旦进入细胞质，一些蛋白质卸下，另一些蛋白质装上，以备蛋白质翻译（第 14 章）。

结合于内含子的 hnRNP 要少，而且组成也不一样。这就是说一组蛋白而不是任何单独一种蛋白质，决定了 RNA 是转运出细胞核还是留在细胞核内。

mRNA 转运是通过核膜上的一种特殊结构完成的，这就是核孔复合体（nuclear pore complex）。分子质量小于 50kDa 左右的小分子可以自由地通过核孔；但是大分子和复合物（包括 mRNA 及其相关蛋白质）则需要通过主动转运。（其余分子，如有些蛋白质，在细胞质中合成但却需要在细胞核内发挥作用，则沿相反的方向转运，即经过同样的核孔从细胞质进入细胞核。）

本书未涉及关于细胞核的跨膜转运机制。我们只需知道，与 RNA 相关的一些蛋白质携带了细胞核跨膜转运信号，转运受体识别出这些信号后指导 RNA 通过核孔离开细胞核。一旦进入细胞质，蛋白质组分就会解离下来，并被识别后重新运回细胞核；然后再与另外的 mRNA 结合，重复下一次循环（图 13-32）。

主动转运需要能量，由 GTP 水解提供，该反应由一种叫做 Ran 的 GTPase 催化。像其他 GTPase 一样，Ran 依据与 GTP 还是 GDP 结合而有两种不同的构象，由一种状态向另一种状态的转变驱动着进出细胞核的转运反应。

小结

大多数基因编码蛋白质，特定蛋白质中的氨基酸序列是由其基因的“密码子”序列决定的。每个密码子由一组 3 个相邻的核苷酸组成。在几乎所有的细菌和噬菌体基因中，可读框都是不间断的一串密码子。但是许多真核生物基因的编码区都被一段段非编码序列割裂成一簇簇密码子。

这些割裂基因中的编码序列叫外显子（“表达序列”），而非编码区域叫内含子（“插入序列”）。某些非编码区也包含在成熟的 mRNA 中—mRNA 的 5′和 3′非翻译区以及整个非编码 RNA 如 microRNA（第 18 章）。因此这些区域也归为外显子。不同基因之间外显子和内含子的数量和大小变化很大。如在酵母中，只有很少一部分基因有内含子，而且即使有，内含子长度短数量也少（每个基因 1 个，偶尔 2 个）。在多细胞生物，如人类中，有内含子的基因数目则大得多，每个基因的内含子数目也很多（最多可达 362 个）。外显子长度确实有所不同，但一般在 150nt 左右；而内含子正好相反，可以从 61bp 到漫长的 800kb 不等。

包含内含子的基因转录时，最初的 RNA 分子中是包括内含子序列的。然后这些内含子被除去，得到成熟的 mRNA。内含子除去的过程叫做剪接。

多数含有内含子的基因只产生一种 mRNA 分子，即每种情况下所有内含子都从初始转录的 RNA 分子中除去，得到的成熟 mRNA 由全部外显子组成。但有时初始 RNA 转录产物可以用不同的形式进行剪接，就会从同一个基因得到许多不同的 mRNA 产物。例如，某些基因含有可供选择的外显子，其中只有一个会留在特定的 mRNA 中。在其他的情况中，一个特定的外显子可能从某些拷贝的 RNA 中去除（与内含子一起）——从而从同一个基因产生不同版本的 mRNA。我们仔细考察了一个选择性剪接的极端例子：果蝇的 *Dscam* 基因。在这个例子中，其中一个外显子有 48 个变异体，并且都出现在 pre-mRNA 上，但只有一个出现在每个成熟的 mRNA 中。对比来自几个物种的基因的序列，通过一流的生物信息学分析揭示了互不相容性选择的原理。

除去内含子是通过两步转酯反应进行的。第一步是由分支点腺苷酸攻击 5′剪接位点的鸟苷酸；第二步是游离的 5′外显子攻击 3′剪接位点。这些反应有两个后果：首先也是最重要的，是把两个外显子融合到了一起；其次，内含子以带分支的套索结构被释放。

对核小体的 pre-mRNA 进行剪接需要一个由多种蛋白质和 RNA 组成的大复合体，即剪接体。剪接体由 snRNP 组成，snRNP 共有 5 种：U1、U2、U4、U5 和 U6 snRNP。每种都由一条 RNA 分子（分别叫做 U1、U2、U4、U5 和 U6 snRNA）和许多蛋白质构成（彼此之间的蛋白质成分大多数不一样）。RNA 成分在识别和催化除去内含子方面起到了核心作用。剪接体组成是动态的，即在剪接反应的不同阶段，剪接体的组成不断改变，不同的组分进进出出，各行使特定的功能。

早期，U1 snRNP 识别 5′剪接位点，而 U2 snRNP 识别分支点。然后 U4 和 U6 结合，与 U5 共同把 5′剪接位点和分支点拉到一起，启动第一步转酯反应，同时 U1 和 U4 离开剪接体。最后，3′和 5′剪接位点被拉到一起，两个外显子连接起来。

有少数内含子能以自剪接的方式把自身从 RNA 分子上除去。虽然算不上真正的酶促反应，但确实是内含子 RNA 介导了剪接的化学反应。自剪接内含子分为两类，其中第Ⅱ类的化学过程与剪接体所介导的相同。这些内含子可能代表了现代由剪接体执行剪接的内含子在进化上的起源，而两者通用的两步转酯反应的化学过程反映了这种进化上的相互关系（也许解释了为何内含子没有采取更直接的一步反应机制进行剪接的原因）。

剪接位点是由很短的、且保守程度不高的序列来确定的。这就给剪接体提出了很大的挑战：如何识别正确的剪接位点并完成剪接？剪接体提高准确性的机制有多种。首先，剪接位点一旦转录出来，剪接体立即就会在相应位置上进行组装，避免了下游位点的竞争。其次，有一些其他蛋白质，如 SR 蛋白，结合在合法位点的附近并协助指引剪接体到达这些位点。这样真实位点与序列类似的假位点相比，肯定对剪接体具有更高的亲和力。

SR 蛋白种类繁多，每一种都是用一个结构域结合 RNA，另一个结构域则和剪接体的某些成分相作用。有些 SR 蛋白调控剪接，表现为某种 SR 蛋白只在一种类型的细胞中出现，从而在这种细胞中完成一种特定的剪接方式。另一些 SR 蛋白只在特定的生理信号下才有活性，所以某种特定的剪接方式仅对应于该生理信号。这样，SR 蛋白类似在后续章节中将会看到的转录激活因子。还与转录调控类似的是，剪接也有抑制因子，可以在某些条件下终止对特定内含子的剪接。我们详细考察了果蝇的性别决定例子，一连串调控的可变剪接事件决定了果蝇是否发育成雄性或雌性。

与第 12 章讨论的其他修饰过程一起，剪接是 mRNA 通过核孔转运出细胞核之前所必须进行的。当然，这也是可调控的。

一般而言一个外显子编码一个独立的折叠或功能蛋白结构域。因此，一个外显子就可以很容易地与其他外显子组合起来发挥作用。这提示，通过进化，比较容易将不同基因中已经存在的外显子重新组合以获得新的蛋白质。

RNA 编辑，是允许 RNA 转录以后进行改变的另一种机制，以使翻译出来的蛋白质与原基因所编码的蛋白质不同。RNA 编辑的机制有两种：酶促碱基修饰和 mRNA 分子内多个尿嘧啶核苷酸的插入或删除。

（刘明旭　宋光　译　吴琳　校）

参考文献

书

Gesteland R.F., Cech T.R., and Atkins J.F., eds. 2006. *The RNA world*, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

剪接和剪接体的作用机制

Crick F. 1979. Split genes and RNA splicing. *Science* **204:** 264–271.

Hastings M.L. and Krainer A.R. 2001. Pre-mRNA splicing in the new millennium. *Curr. Opin. Cell Biol.* **13:** 302–309.

Hertel K.J. and Graveley B.R. 2005. RS domains contact the pre-mRNA throughout spliceosome assembly. *Trends Biochem. Sci.* **30:** 115–118.

Maniatis T. and Reed R. 2002. An extensive network of coupling among gene expression machines. *Nature* **416:** 499–506.

Staley J.P. and Guthrie C. 1998. Mechanical devices of the spliceosome: Motors, clocks, springs, and things. *Cell* **92:** 315–326.

Tange T.Ø., Nott A., and Moore M.J. 2004. The ever-increasing complexities of the exon junction complex. *Curr. Opin. Cell Biol.* **16:** 279–284.

Tarn W.Y. and Steitz J.A. 1997. Pre-mRNA splicing: The discovery of a new spliceosome doubles the challenge. *Trends Biochem. Sci.* **22:** 132–137.

自我剪接

Cech T.R. 1990. Nobel lecture. Self-splicing and enzymatic activity of an intervening sequence RNA from *Tetrahymena*. *Biosci. Rep.* **10:** 239–261.

替换剪接及调控

Blencowe B.J. 2006. Alternative splicing: New insights from global analyses. *Cell* **126:** 37–47.

Graveley B.R. 2001. Alternative splicing: Increasing diversity in the proteomic world. *Trends Genet.* **17:** 100–107.

Hughes T.A. 2006. Regulation of gene expression by alternative untranslated regions. *Trends Genet.* **22:** 119–122.

Kornblihtt A.R. 2005. Promoter usage and alternative splicing. *Curr. Opin. Cell Biol.* **17:** 262–268.

Ladd A.N. and Cooper T.A. 2002. Finding signals that regulate alternative splicing in the post-genomic era. *Genome Biol.* **3:** reviews0008.1-0008.16.

Maniatis T. and Tasic B. 2002. Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature* **418:** 236–243.

Smith C.W.J. 2005. Alternative splicing—When two's a crowd. *Cell* **123:** 1–3.

Xing Y. and Lee C. 2006. Alternative splicing and RNA selection pressure—Evolutionary consequences for eukaryotic genomes. *Nat. Rev. Genet.* **7:** 499–509.

mRNA转运

Dreyfuss G., Kim V.N., and Kataoka N. 2002. Messenger-RNA-binding proteins and the messages they carry. *Nat. Rev. Mol. Cell Biol.* **3:** 195–205.

进化

Roy S.W. and Gilbert W. 2006. The evolution of spliceosomal introns: Patterns, puzzles and progress. *Nat. Rev. Genet.* **7:** 211–221.

RNA编辑

Blanc V. and Davidson N.O. 2003. C-to-U RNA editing: Mechanisms leading to genetic diversity. *J. Biol. Chem.* **278:** 1395–1398.

Chiu Y.-L. and Greene W.C. 2006. Multifaceted antiviral actions of APOBEC3 cytidine deaminases. *Trends Immunol.* **27:** 291–297.

———. 2006. APOBEC3 cytidine deaminases: Distinct antiviral actions along the retroviral life cycle. *J. Biol. Chem.* **281:** 8309–8312.

Decatur W.A. and Fournier M.J. 2003. RNA-guided nucleotide modification of ribosomal and other RNAs. *J. Biol. Chem.* **278:** 695–698.

Keegan L.P., Leroy A., Sproul D., and O'Connell M.A. 2004. Adenosine deaminases acting on RNA (ADARs): RNA-editing enzymes. *Genome Biol.* **5:** 209.

Madison-Antenucci S., Grams J., and Hajduk S.L. 2002. Editing machines: The complexities of trypanosome RNA editing. *Cell* **108:** 435–438.

Sattelle D.B., Jones A.K., Sattelle B.M., Matsuda K., Reenan R., and Biggin P.C. 2005. Edit, cut and paste in the nicotinic acetylcholine receptor gene family of *Drosophila melanogaster*. *BioEssays* **27:** 366–376.

Simpson L., Sbicego S., and Aphasizhev R. 2003. Uridine insertion/deletion RNA editing in trypanosome mitochondria: A complex business. *RNA* **9:** 265–276.

Stuart K.D., Schnaufer A., Ernst N.L., and Panigrahi A.K. 2004. Complex management: RNA editing in trypanosomes. *Trends Biochem. Sci.* **30:** 97–105.

Turelli P. and Trono D. 2005. Editing at the crossroad of innate and adaptive immunity. *Science* **307:** 1061–1065.

剪接与疾病

Faustino N.A. and Cooper T.A. 2003. Pre-mRNA splicing and human disease. *Genes Dev.* **17:** 419–437.

Licatalosi D.D. and Darnell R.B. Splicing regulation in neurologic disease. *Neuron* **52:** 93–101.

Srebrow A. and Kornblihtt A.R. 2006. The connection between splicing and cancer. *J. Cell Sci.* **119:** 2635–2641.

Yeo G.W.-M. 2005. Splicing regulators: Targets and drugs. *Genome Biol.* **6:** 240.

第14章 翻译

本章概要

•

信使 RNA

•

转移 RNA

•

连接氨基酸到 tRNA 上

•

核糖体

•

翻译的起始

•

翻译延伸

•

翻译终止

•

翻译的调控

•

依赖翻译的 mRNA 和蛋白质的稳定性调节

本章和下一章讨论的中心问题是蕴藏在信使 RNA（messenger RNA）核苷酸序列中的遗传信息是如何被解读产生蛋白质氨基酸顺序的，这一过程称作翻译（translation）。在我们所讨论的所有事件中，翻译是所有生物中最保守，也是细胞中最耗费能量的事件之一。在快速生长的细菌细胞中，最多可有 80%的能量和 50%的细胞干重是专门用于蛋白质合成的。实际上，单一蛋白质的合成需要上百种蛋白质和 RNA 的合作。与翻译过程较为复杂的本质相一致，我们分两章对其进行讨论。本章描述解码mRNA的一系列事件，第 15 章描述遗传密码的本质和转运 RNA 对其的识别。

在信息传递中，翻译远比 DNA 到 RNA 的转录复杂。与 DNA 模板和信使 RNA 的核苷酸的互补性不同的是，氨基酸的侧链与 RNA 的嘧啶和嘌呤之间简直没有什么特异的亲和力。例如，丙氨酸（Ala）、缬氨酸（Val）、亮氨酸（Leu）和异亮氨酸（Ile）的疏水侧链不能和核苷酸碱基的氨基和酮基形成疏水键。同样地，很难想像 RNA 的三个碱基的不同组合能够形成与芳香族氨基酸苯丙氨酸（Phe）、酪氨酸（Tyr）和色氨酸（Trp）特异亲和的表面。因此，多肽链中特异和准确的氨基酸顺序似乎不可能是由 mRNA 模板和氨基酸的直接作用导致的。

考虑到上述问题，1955 年 Francis H. Crick 提出在形成多肽链之前，氨基酸必须结合一个能够直接识别并作用于信使 RNA 的三核苷酸密码子的特别的转配器（adaptor）分子。考虑到需要依据 Watson-Crick 法则来识别密码，Crick 猜测这一转配器可能是 RNA 分子。仅仅两年之后，Paul C. Zamecnik 和 Mahlon B. Hoagland 就证明了在蛋白

质形成之前，氨基酸确实和一组 RNA 分子（约占所有细胞 RNA 的 15%）相结合。这些 RNA 称作转运 RNA（或 tRNA），因为它们携带的氨基酸随后被转运到不断增长的多肽链上。

将信使 RNA 的语言翻译成蛋白质语言的机器由 4 种基本的成分组成：信使 RNA（mRNA）、转移 RNA（tRNA）、氨基酰 tRNA 合成酶（aminoacyl tRNA synthetase）和核糖体（ribosome）。通过合作，它们完成了一项非凡的任务：将一种由 4 个字母书写的密码翻译成由 20 个氨基酸书写的密码。作为翻译的模板，mRNA 提供了必须由翻译机器来解读的信息，mRNA 的蛋白质编码区是由称作密码子（codon）的一系列有序的三核苷酸单位组成的，密码子指定氨基酸的顺序。tRNA 提供了将要加入多肽链的氨基酸和 mRNA 的密码子之间的物理表面。氨基酰 tRNA 合成酶使氨基酸和识别正确密码子的 tRNA 结合起来。翻译过程最后一个重要角色是核糖体，它是一个由 RNA 和蛋白质组成的分子质量为数百万道尔顿的庞大的细胞器。核糖体协调着 tRNA 对 mRNA 的正确识别，并催化形成介于正在合成的多肽链和已结合到特定 tRNA 上的氨基酸之间的肽键。

我们将首先讨论 4 个组分中每一组分的主要特征，然后描述这 4 种组分如何共同作用以完成翻译。近来对翻译机器各组分的结构研究的进展使得翻译过程成为激动人心的研究领域——充满着机械论洞察力的领域。我们将要提出以下的问题：mRNA 中核苷酸序列信息的组织结构是什么？tRNA 的结构是什么？氨基酰 tRNA 合成酶是如何识别并结合正确的氨基酸到各个 tRNA 上的？最后，核糖体是如何组织解码核苷酸序列信息和把氨基酸加入到正在增长的多肽链上的？

信使 RNA

多肽链是由可读框规定的

翻译机器只解码每条 mRNA 的部分序列。如在第 2 章中所看到的及在第 15 章中将要详细讨论的，蛋白质合成的信息蕴藏在三联核苷酸密码子中，每个密码子指定一个氨基酸。每条 mRNA 的蛋白质编码区由连续的、不交叠的、称作可读框（open-reading frame，ORF）的密码子串组成。每个 ORF 对应一个蛋白质，其起始和终止点都位于 mRNA 内部，即 ORF 的终止点不同于 mRNA 的终止点。

翻译始于可读框的 5′端起点，一个密码子一个密码子地通向 3′端终点。第一个和最后一个密码子分别称作起始密码子（start codon）和终止密码子（stop codon）。细菌中，起始密码子通常是 5′-AUG-3′，但是 5′-GUG-3′有时甚至 5′-UUG-3′也使用。真核细胞总是使用 5′-AUG-3′为起始密码子。起始密码子有两个重要的功能，第一，它指定掺入到增长的多肽链中的第一个氨基酸；第二，它确定了所有后续密码子的可读框。由于密码子是一个紧接一个（但不交叠），而且由三个核苷酸组成，因此，每条 mRNA 链可以用 3 种不同的可读框来翻译（图 14-1）。但是，一旦翻译开始了，可读框就确定

了。这样，通过固定第一个密码子的位置，起始密码子决定了所有后续密码子的位置。

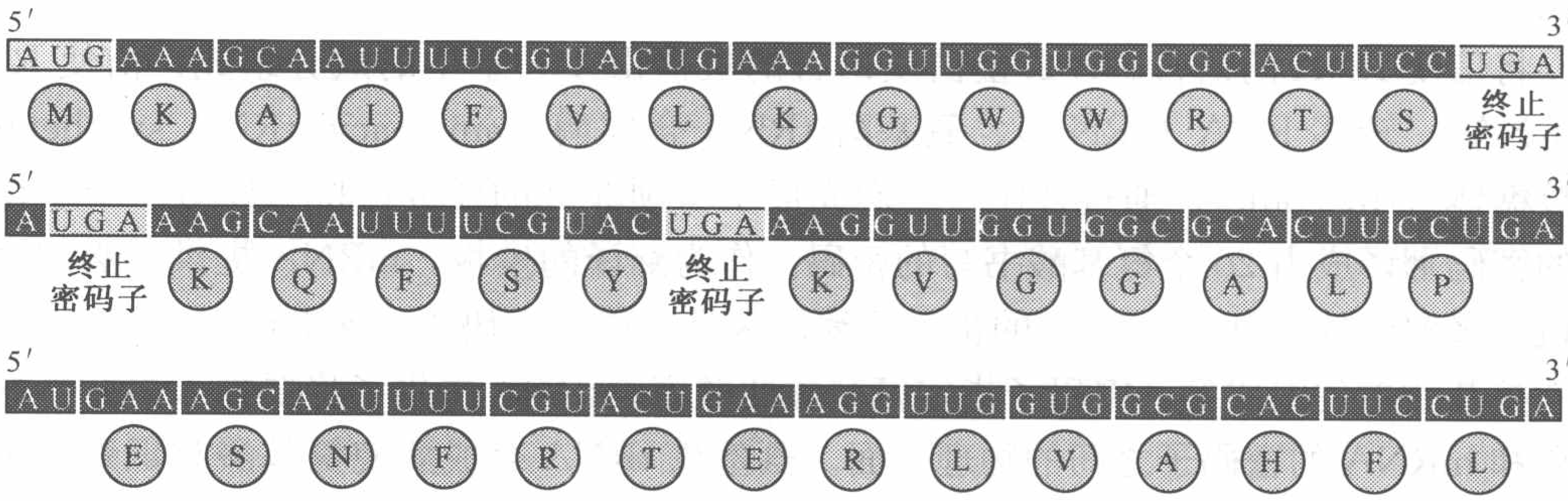


图 14-1 大肠杆菌 *Trp* 前导序列的 3 种可能的可读框。起始密码子以绿色表示，终止密码子以红色表示。每个可读框的氨基酸的序列在每一密码子的下方以字母表示。

终止密码子有 3 个（5′-UAG-3′、5′-UGA-3′、5′-UAA-3′），它们决定可读框的终点并发出终止多肽合成的信号。我们现在可以全面地理解可读框这一术语的含义了。它是一段连续的密码子，以特定的框架（由第一个密码子决定）来“阅读”，该框架对翻译是“开放”的，因为它没有终止密码子（也就是说，直到 ORF 的最后一个密码子）。

信使 RNA 包含至少一个可读框。每条 mRNA 的 ORF 数量在真核和原核细胞中是有区别的。真核细胞的 mRNA 几乎都只有一个 ORF。相反，原核细胞的 mRNA 经常含有两个或多个 ORF，因此可编码多个多肽链。含有多个 ORF 的信使 RNA 称作多顺反子 mRNA（polycistronic mRNA），那些只编码一个 ORF 的信使 RNA 称作单顺反子 mRNA（monocistronic mRNA）。正如在第 12 章中所学到的，细菌中见到的多顺反子 mRNA 经常编码功能相关的蛋白质，如在氨基酸或核酸的生物合成过程中参与不同步骤的蛋白质。典型的原核和真核细胞的 mRNA 结构如图 14-2 所示。

原核细胞 mRNA 具有核糖体结合位点，用以募集翻译机器

要使翻译发生，mRNA 必须募集核糖体。为了方便与核糖体结合，许多原核细胞的可读框在起始密码子的上游（5′端）含有一小段称为核糖体结合位点（ribosome binding site，RBS）的序列。这一元件也被称作 SD 序列（Shine-Dalgarno sequence），是以比较了多个 mRNA 序列后发现它的科学家的名字命名的。RBS 典型地位于起始密码子 5′端 3～9 个碱基内，与核糖体 RNA 组分之一的 16S 核糖体 RNA（rRNA）（图 14-2a）的 3′端的一段序列互补。RBS 与这一 RNA 组分发生碱基配对，使得核糖体排列在可读框的起始处。16S RNA 的这一区域的核心序列为5′-CCUCCU-3′。无独有偶，原核细胞的 RBS 大多具有 5′-AGGAGG-3′序列。RBS 和起始密码子之间碱基的互补状况和空间距离对可读框翻译的活跃程度有很大的影响：互补度高并且距离适合能提高翻译的水平；反之互补度低和/或距离不当则通常只有较低水平的翻译。

某些原核细胞的 ORF 缺乏较强的 RBS 但仍有活跃的翻译。这些 ORF 不是 mRNA

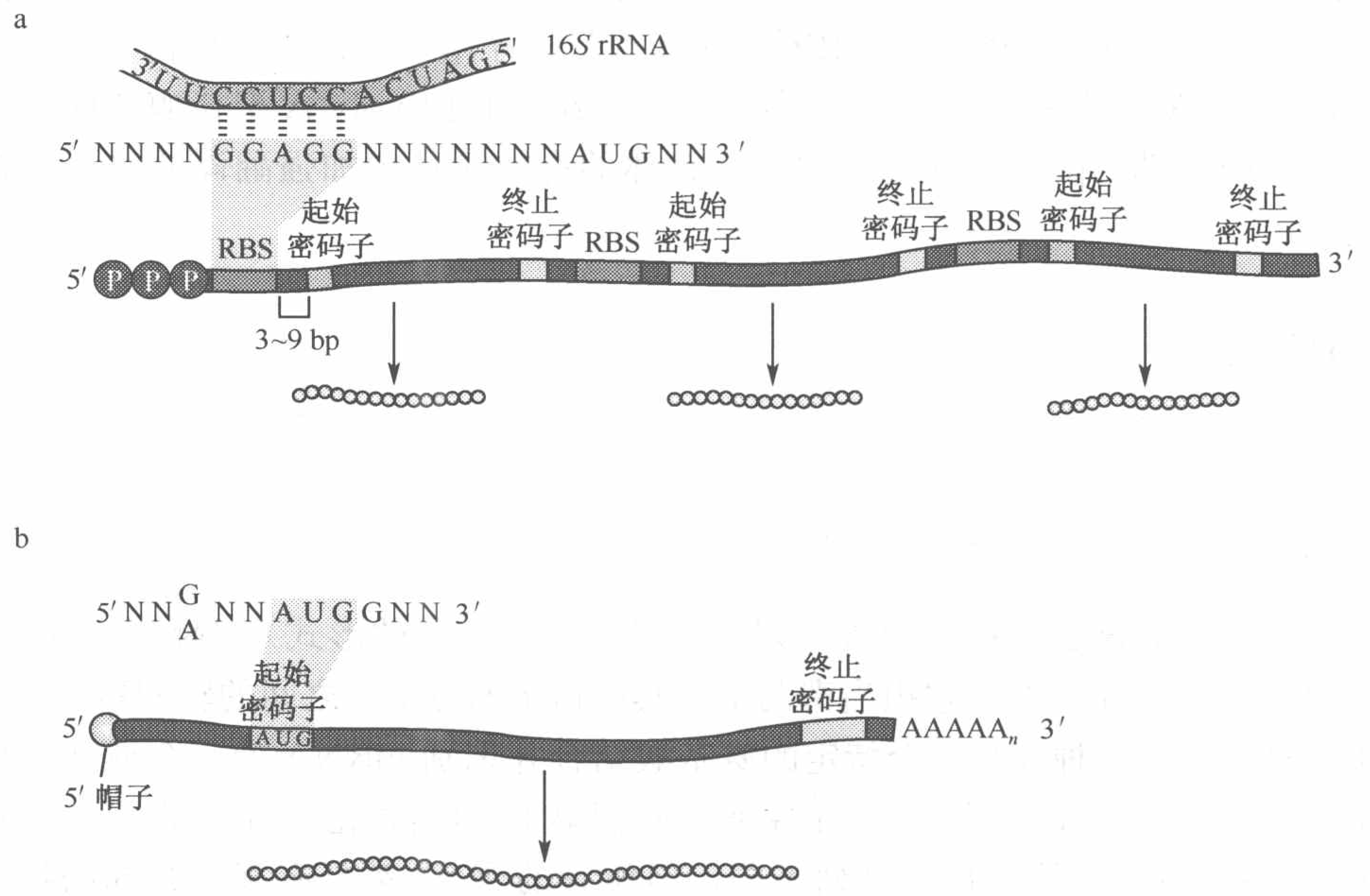


图 14-2 信使 RNA 的结构。a. 一个多顺反子的原核细胞信使 RNA。核糖体结合位点以 RBS 表示。b. 一个单顺反子的真核信使 RNA。5′帽子结构在 mRNA 的末端以球形表示。

的第一个 ORF，而位于多顺反子信使 RNA 中另一个 ORF 之后（不是所有的原核 mRNA 是多顺反子）在这些情况下，下游可读框的起始密码子经常与相邻的可读框 3′端相交叠（大多数是5′-AUGA-3′序列，既有起始密码子，也有终止密码子）。这样，刚刚结束上游可读框的翻译的核糖体正好处于恰当的位置，并从下游可读框的起始密码子开始翻译，这种结构安排绕开了用 RBS 募集核糖体的需要。这种交叠可读框之间的连锁翻译称为翻译耦合（translational coupling）。需要注意的是，在这种情况下，下游 ORF 的翻译需要上游 ORF 的翻译。事实上，对于两个翻译耦合的基因，上游 ORF 中的终止突变也会阻止下游 ORF 的翻译。

真核细胞 mRNA 在 5′和 3′端被修饰，以促进翻译

与原核细胞不同，真核细胞的 mRNA 通过位于 5′末端的称为 5′帽子（5′cap）的特殊的化学修饰来募集核糖体（第 12 章和图 14-2b）。所谓 5′帽子是一个通过与众不同的 5′-5′结合方式与 mRNA 的 5′端结合的甲基化的鸟嘌呤核苷酸。经过 3 个步骤（第 12 章图 12-23），5′帽子的鸟嘌呤核苷酸与 mRNA 的 5′端通过 3 个磷酸基团结合，这样形成的结构吸引核糖体到 mRNA 上。一旦结合到 mRNA 上，核糖体沿着 5′→3′的方向运行直至遇到5′-AUG-3′起始密码子，这一过程称作扫描（scanning）。

真核生物 mRNA 有另外两个特征利于翻译。其一是在某些 mRNA 中，起始密码子的上游第三个碱基为嘌呤并且第一个下游碱基为鸟嘌呤（5′-G/ANNAUGG-3′）。这一序列首先由 Marilyn Kozak 发现，称为 Kozak 序列。虽然许多真核细胞缺乏 Kozak 序

列，但它的存在确实能够提高翻译的效率。与原核细胞的情况相反，Kozak 序列据认为与起始 tRNA 相互作用，而不是与核糖体的 RNA 成分作用。另一个有利于翻译的特征是 3′端的 poly-A 尾结构。如第 12 章所示，这一结构是由 poly-A 聚合酶合成的。尽管位于mRNA的 3′端，poly-A 可通过提高核糖体的有效循环来加强翻译的水平（将在后面讨论）。

转移 RNA

tRNA 是密码子和氨基酸之间的转配器

蛋白质合成的核心是将核苷酸序列的信息（以密码子的形式）“翻译”成氨基酸。这是由 tRNA 分子完成的，它担当密码子及其所指定的氨基酸之间的转配器。tRNA 分子有很多种，但每一种仅与一个特定的氨基酸结合并识别 mRNA 的一个或几个特定的密码子（大多数 tRNA 识别一个以上密码子我们第 15 章将讨论）。tRNA 分子的长度为 75～95 个核苷酸。虽然各自的序列不同，但所有的tRNA都具有某些共同的特征。首先，所有的 tRNA 在 3′端均以 5′-CCA-3′序列结束。(框 14-1）与该保守性相一致的是，该序列（和 tRNA 的）3′端是与其编码的氨基酸相结合的位点。

tRNA 另一个显著的特征就是在其一级结构中存在一些特殊的碱基，这些特殊的碱基是多核苷酸链的正常碱基在转录后由酶的修饰作用形成的。例如，假尿嘧啶（ΨU）就是尿苷经异构化作用使尿嘧啶与核糖结合的位置从环 1 位的 N 转移到环 5 位的 C 而形成的(图 14-3)。同样地，双氢尿嘧啶（dihydrouridine，D）是尿苷经酶的作用使 5 位和 6 位之间的 C 双键减为单键而形成的。tRNA 中发现的另一些特殊碱基包括次黄嘌呤(hypoxanthine，H)、胸腺嘧啶（thymine，T）和甲基鸟嘌呤（methylguanine，G^m）。这些修饰碱基对 tRNA 的功能并非不可或缺，但缺少这些碱基的细胞生长变慢。这一现象提示这些修饰碱基可能提高 tRNA 的功能。例如，如第 15 章将要看到的，次黄嘌呤在某些 tRNA 对密码子的识别过程中扮演重要的角色。

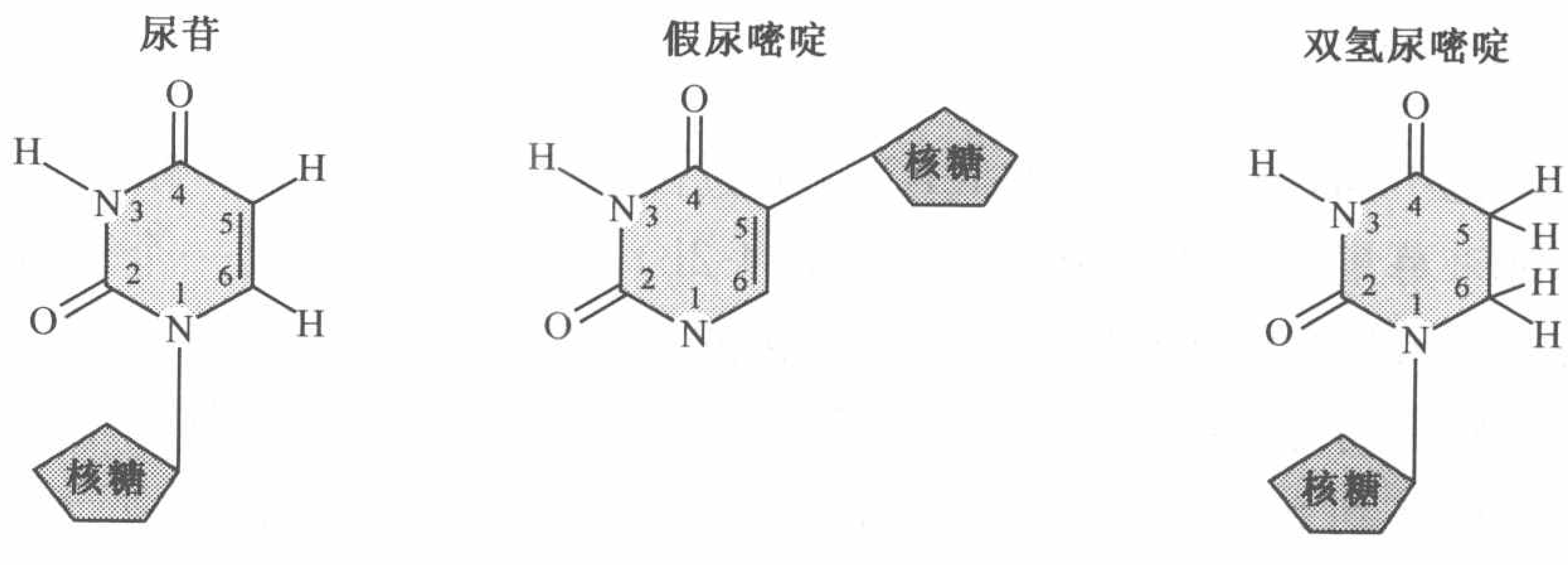


图 14-3　tRNA 中发现的一组修饰核苷。图示尿苷和 2 个尿苷相关的核苷。

框 14-1 CCA 添加（CCA-Adding）酶：无模板合成 RNA

如正文所述，5′-CCA-3′端在所有 tRNA 中是保守的，并且在蛋白质合成中是必需的。但奇怪的是，当克隆那些编码 tRNA 的基因时，却发现许多并不编码 CCA 端。反而这些基因比成熟的 tRNA 3′端短 3 个核苷酸。事实上，许多编码细菌 tRNA 和几乎所有真核生物 tRNA 的基因都缺乏末端的这 3 个碱基。那么，这些 tRNA 是怎样获得其 CCA 端的呢？答案来自一种称为 CCA-添加酶的特殊 RNA 合成酶。顾名思义，这种酶可以将 CCA 添加到末端原本缺乏这个序列的 tRNA 上。奇怪的是，这种酶没有对应的核酸组分；也就是说，CCA-添加酶在没有 RNA 或 DNA 模板的情况下将一段特定的序列添加到 tRNA 的末尾。

CCA 添加酶如何在没有模板的情况下加入一段特定的序列呢？对这些酶的一系列三维结构开始揭示其解决方案。首先，像其他 RNA 合成酶和 DNA 合成酶一样，CCA 添加酶只有一个活性位点，采用类似的双金属离子催化机制（图 8-6）。在这一活性位点内部，一个氨基酸和来自 tRNA 末尾核苷酸的一个磷酸基团与 A 和 C 而不与 G 和 U 形成氢键。可以通过观察每个碱基的氢键供体和受体的键型来理解这种特异性（图 6-5 及图 6-6）。A 和 C 的键型有重叠，而 G 和 U（或 T）相反和互补。实际上，正是这种互补性担负着 DNA 双螺旋内的碱基配对的特异性。这种氢键键型解释了 CCA 添加酶对 A 和 C 的特异性。

添加 C 还是 A 的特异性是由加入碱基时酶活性位点的变化来控制的。与其他聚合酶不同，当核苷加入时 tRNA 模板并不改变位置。相反，tRNA 模板紧紧地固定在其位置上，每个加入的核苷改变了酶活性位点的结构。改变的结果是，当缺失 CCA 的 tRNA 与酶结合时，酶活性位点对 C 具有特异性，而当加入了两个 C 之后，酶变得对 A 特异性。一旦 A 加入之后，活性位点不再对任何多余的碱基开放，而带有新加入的 CCA 末端的 tRNA 得以释放。

tRNA 都有三叶草型的二级结构

在第 6 章中曾谈到，RNA 分子通常含有自我互补区域，能够通过碱基配对形成有限的双螺旋结构。RNA 的其他区域因没有互补结构，所以是单链的。tRNA 分子表现为特异的高度保守的单链和双链区域（二级结构），该结构可描绘成三叶草型结构（图 14-4）。tRNA 三叶草型结构主要特征是有一受体臂、3 个茎环（分别为 ΨU 环、D 环和反密码子环）和第四个可变环。对于这些特征结构的描述如下：

- 受体臂因是结合氨基酸的位点而得名，由 tRNA 分子的 5′和 3′端的碱基配对而成。其 3′端的5′-CCA-3′序列伸出双链之外。
- ΨU 环因特殊碱基 ΨU（假尿嘧啶）的存在而得名。这一修饰碱基经常发现位于 5′-TΨUCG-3′序列中。
- D 环因特征性的双氢尿嘧啶的存在而得名。
- 反密码子环，顾名思义，包含反密码子，一个通过碱基配对识别 mRNA 的密码子的三核苷酸序列。反密码子的两端由 5′端的尿嘧啶和 3′端的嘌呤界定。

• 可变环位于反密码子环和 ΨU 环之间，长度从 3～21bp 不等，正如其名所示。

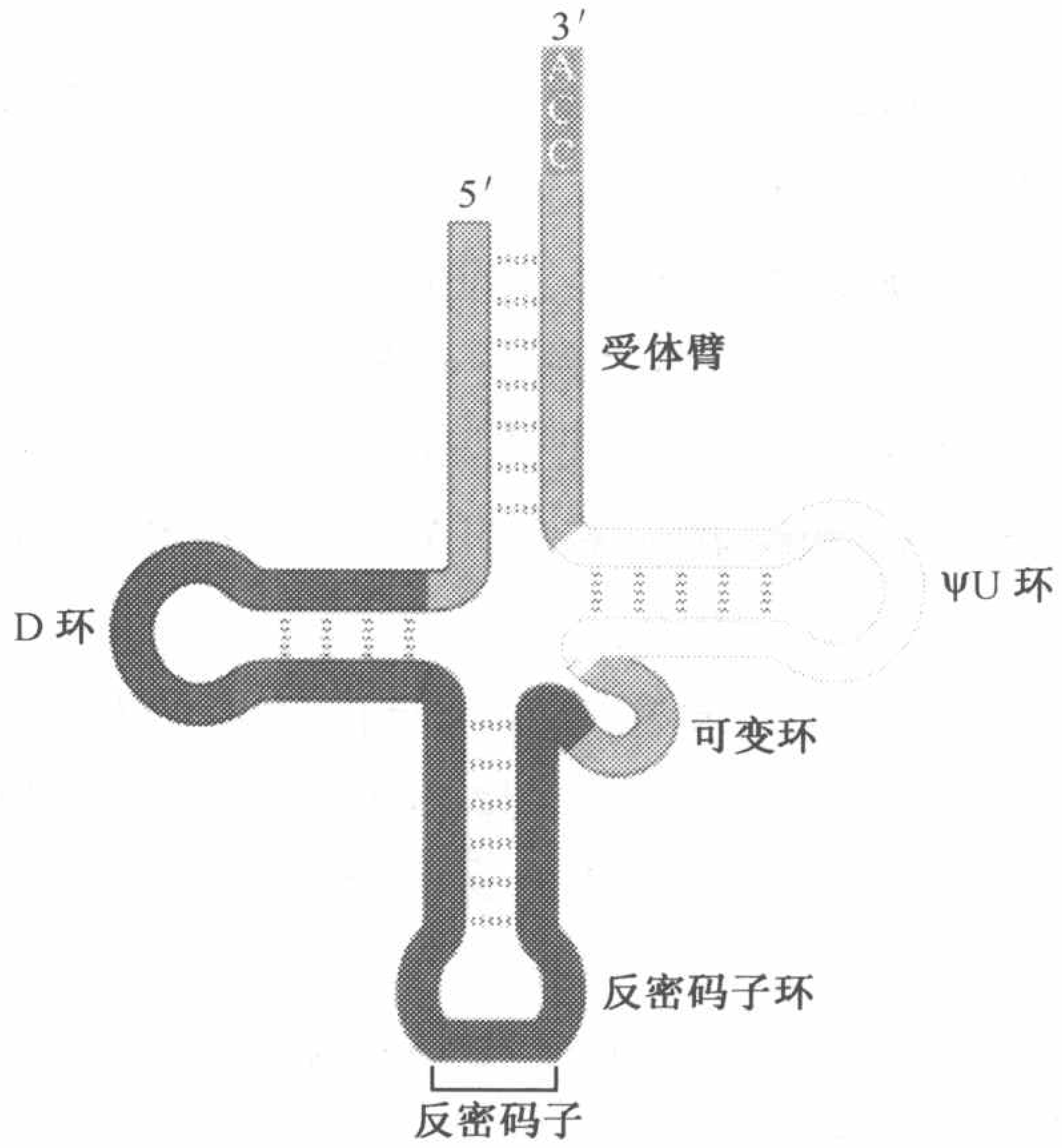


图 14-4 tRNA 的三叶草状二级结构。在此 tRNA 示意图中，红色虚线表示 tRNA 不同部分之间的碱基配对。

tRNA 具有 L 状的三维结构

三叶草型结构揭示了 tRNA 含有自我互补的区域。那么，这一转配器分子的实际三维构象如何呢？X 射线衍射揭示其为 L 状结构，受体臂在一端，反密码子环在另一端，两者相距 70Å（图 14-5c)。要了解这 L 状结构和三叶草型结构的关系，请考虑以下方面：在 tRNA 的三维构象中，受体臂和 ΨU 环的茎之间形成延伸的螺旋结构（图 14-5b)。同样地，反密码子茎和 D 环的茎形成第二个延伸的螺旋。这两个延伸的螺旋垂直排列，D 环和 ΨU 环靠在一起。

三种相互作用使这 L 状结构得以稳定。第一，因形成两个延伸的碱基配对区域而获得的类似 DNA 双链的碱基堆积作用。第二，三维结构中相互靠近的不同螺旋区域的碱基间的氢键，这些通常是非常规（非 Watson-Crick）键。第三，碱基和磷酸核糖骨架之间的相互作用。

连接氨基酸到 tRNA 上

氨基酸通过高能酰基连接到 tRNA 3′端的腺苷酸上使 tRNA 负载

连接了氨基酸的 tRNA 分子称为负载的（charged）tRNA，未连接氨基酸的 tRNA 称为空载的（uncharged）tRNA。这一负载过程需要氨基酸的羧基与 tRNA3′端受体臂

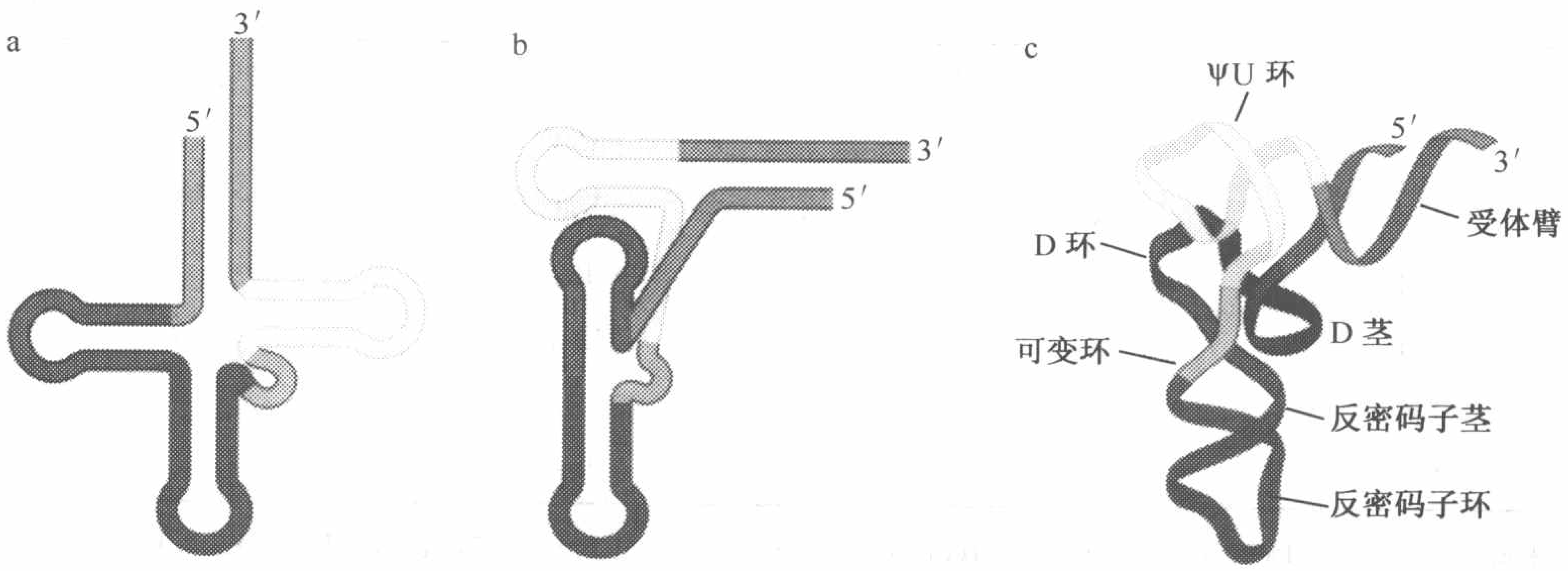


图 14-5 tRNA 的三叶草结构和实际的三维结构之间的转换。 a. 三叶草示意图；b. L 状结构示意图代表最终折叠的 tRNA 碱基配对的位置；c. tRNA 的实际折叠结构的带状示意图。请注意虽然此图说明了实际的 tRNA 结构与三叶草示意图的关系，但 tRNA 并不是通过先碱基配对再 L 状折叠来形成最后的结构的。

突出的腺苷酸的 2′或 3′羟基（见下）形成酰基。形成的酰基是一个高能键，因为它的水解导致自由能的较大变化。这对于蛋白质的合成是非常重要的：这一高能键水解所释放的能量帮助推动多肽链中氨基酸之间肽键的形成。

氨基酰 tRNA 合成酶分两步使 tRNA 负载

所有的氨基酰 tRNA 合成酶分两个酶促步骤使氨基酸连接到 tRNA 上（图 14-6）。第一步是腺苷酰基化（adenylylation），氨基酸和 ATP 反应而被腺苷酰基化，并同时释放出焦磷酸。腺苷酰基化指的是 AMP 的转移，与腺苷酸化（adenylation）相反，后者发生的是腺苷酸的转移。正如我们在多核苷酸的合成过程中看到的一样（第 8 章），腺苷酰基化反应的主要驱动能量来自于随后的焦磷酸酶对焦磷酸的水解。腺苷酰基化的结果是，通过氨基酸的羧基与 AMP 的磷酸基团形成的高能酯键，氨基酸与腺苷酸结合。第二步是 tRNA 负载（tRNA charging），与氨基酰 tRNA 合成酶保持紧密结合的腺苷酰基化的氨基酸与 tRNA 反应。这一反应导致氨基酸通过 2′或 3′羟基被转移到 tRNA 的 3′端，同时释放出 AMP。

tRNA 合成酶分两类（表 14-1）。Ⅰ类酶将氨基酸连接到 tRNA 的 2′羟基，并且通常是单聚体；Ⅱ类酶将氨基酸连接到 tRNA 的 3′羟基，并且通常是二聚体或四聚体。虽然氨基酸与 tRNA 之间最初的耦连方式不同，但一旦从合成酶中释放出来，氨基酸能够很快地平衡其与 2′或 3′羟基的连接。

表 14-1 氨基酰 tRNA 合成酶的种类*

Ⅱ类	四级结构	Ⅰ类	四级结构
Gly	($\alpha_2\beta_2$)	Glu	(α)
Ala	(α_4)	Gln	(α)
Pro	(α_2)	Arg	(α)
Ser	(α_2)	Cys	(α_2)

续表

Ⅱ类	四级结构	Ⅰ类	四级结构
Thr	(α_2)	Met	(α_2)
His	(α_2)	Val	(α)
Asp	(α_2)	Ile	(α)
Asn	(α_2)	Leu	(α)
Lys	(α_2)	Tyr	(α)
Phe	($\alpha_2\beta_2$)	Trp	(α)

来源：Data from Delarue M. 1995. Current Opinion in Structural Biology 5：48-55. Table 1. © Elsevier.

＊Ⅰ类酶一般是单体的，而Ⅱ类酶是二聚体或四聚体的，来自两个亚基的残基为单一的 tRNA 提供结合位点。α 和 β 指 tRNA 合成酶的亚基，下标指化学计量。

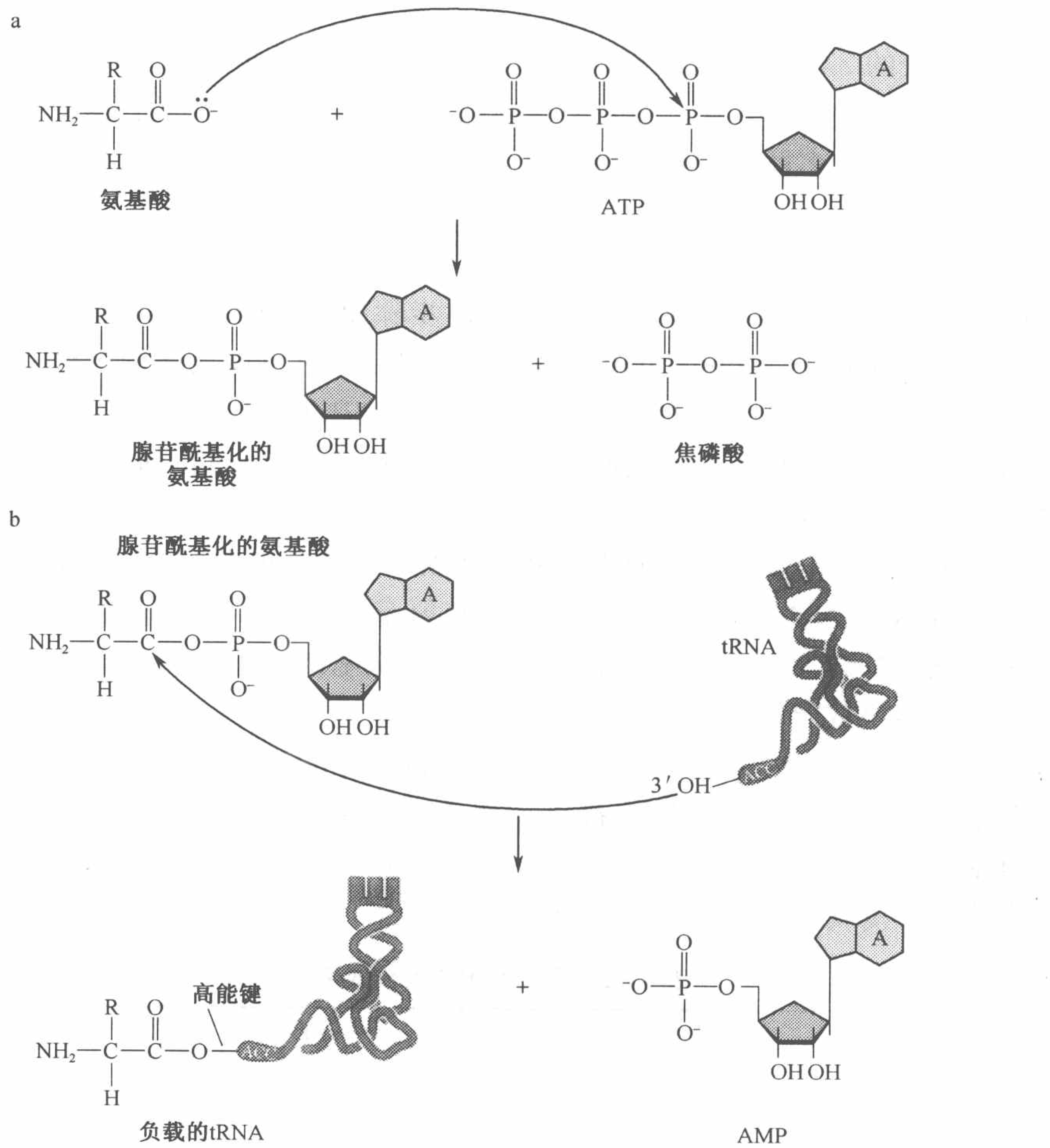


图 14-6　氨基酰 tRNA 负载的两个步骤。 a. 氨基酸的腺苷酰基化；b. 腺苷酰基化的氨基酸转移至 tRNA 上。此图显示的为Ⅱ类 tRNA 合成酶的反应（该酶连接氨基酸到 3′-OH 上）

每种氨基酰 tRNA 合成酶连接一个氨基酸到一个或多个 tRNA 上

20 种氨基酸的每一种都是通过单种专门的 tRNA 合成酶连接到正确的 tRNA 上的。由于大多数的氨基酸是由多于一个的密码子决定的（第 15 章），因此一个合成酶经常识别多于一个的 tRNA（称为同工 tRNA，isoaccepting tRNA）并使之负载氨基酸。但是，同一种 tRNA 合成酶需要负责一个特定氨基酸的所有不同的 tRNA 的负载。因此，对一个氨基酸而言，有一种，而且只有一种 tRNA 合成酶将其连接到所有正确的 tRNA 上。

大多数的生物具有 20 种不同的 tRNA 合成酶，但并不总是这样。例如，有些细菌缺乏使谷酰胺连接到 $tRNA^{Gln}$的合成酶。相反，由一种氨基酰 tRNA 合成酶负责将 Glu 连接到 $tRNA^{Gln}$和 $tRNA^{Glu}$上。然后第二种酶将负载的 $tRNA^{Glu}$上的 Glu 氨基化形成 Gln，即将 Glu-$tRNA^{Gln}$氨基化变成 Gln-$tRNA^{Gln}$（前缀指氨基酸，上标指 tRNA 所识别的密码子）。第二种酶的存在使细菌不再需要谷酰胺 tRNA 合成酶。但是，一种氨基酰 tRNA 合成酶永远不能将多于一种的氨基酸连接到一种特定的 tRNA 上。

tRNA 合成酶识别同族 tRNA（cognate tRNA）的独特结构

从前面的讨论中可以看出，氨基酰 tRNA 合成酶面临两个重要的挑战：它们必须识别针对特定氨基酸的一组正确的 tRNA，而且它们必须使这些同工 tRNA 负载正确的氨基酸。这两个过程都必须高度精确地进行。

让我们首先探讨 tRNA 识别的特异性：tRNA 分子的什么特征使得合成酶能够将某种氨基酸的同族的同工 tRNA 分子同其他 19 种氨基酸的 tRNA 分子区别开呢？来自遗传学、生物化学和 X 射线衍射的证据表明决定这特异性的因素来自 tRNA 分子的两个

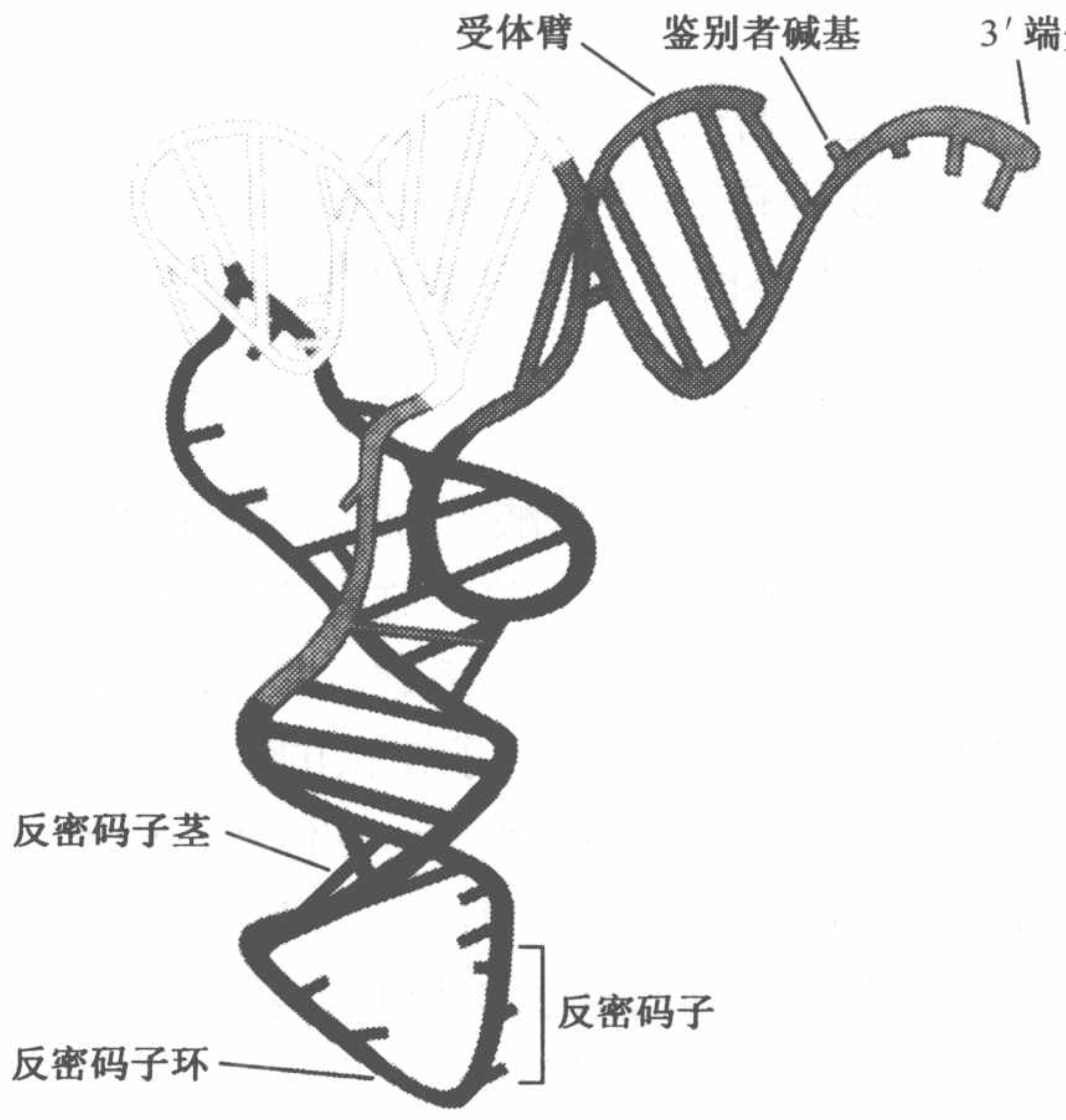


图 14-7　tRNA 的结构：氨基酰 tRNA 合成酶识别的必需元件。



图 14-8 谷氨酰 tRNA 合成酶与 $tRNA^{Gln}$ 结合的晶体结构图。酶以灰色表示，$tRNA^{Gln}$ 以紫红色表示，黄、红和绿色的分子是谷氨酰-AMP。请注意这一分子和 $tRNA^{Gln}$ 的 3′端的接近性以及 tRNA 和合成酶作用的结合点。（来源：Rath V. L. et al. 1998. Structure 6: 439-449.）图像使用 BobScript、MolScript 和 Raster 3D 软件制作。

相距较远的位点：受体臂和反密码子环（图 14-7）。受体臂是 tRNA 合成酶识别特异性的一个极其重要的决定因素。在某些情况下改变受体臂的一对碱基［称为鉴别者碱基（discriminator base)］足以使某个 tRNA 的识别特异性从一种合成酶变为另一种合成酶。虽然如此，反密码子环也经常对识别特异性有所贡献。例如，谷酰胺 tRNA 合成酶同受体臂和反密码子环都有广泛的接触，包括反密码子本身（图 14-8）。

你也许会认为由于反密码子是 tRNA 的最终决定特征——反密码子决定了 tRNA 负责加入到多肽链上的氨基酸种类，因此反密码子在 tRNA 合成酶的识别过程中总是起作用。但是，由于每个氨基酸通常由多于一个的密码子所编码，因此在很多情况下对反密码子的识别作用并不起作用。例如，Ser 由 6 个密码子编码，包括 5′-AGC-3′ 和 5′-UCA-3′这两个完全不同的密码子。这样，Ser 的 tRNA 需要具有不同的反密码子，而无法轻易地被单一的 tRNA 合成酶所识别。为此，Ser 的合成酶必须依靠反密码子以外的决定区来识别 tRNA 分子。

因为它们在信息传递过程中具有重要作用，这些使合成酶鉴别不同 tRNA 的 tRNA 分子上的决定结构有时被称为“第二遗传密码”。正如以上所讨论的，这一套密码远比“第一遗传密码”复杂，并且不易列成表格。然而，没有这一套密码，合成酶不能将不同的 tRNA 分子区别开来，翻译体系也不能合成具有重复序列的多肽链。

氨基酰 tRNA 的形成是非常精确的

氨基酰 tRNA 合成酶在选择正确的氨基酸方面所面临的挑战甚至比其在识别正确的 tRNA 分子上的挑战更为令人吃惊（图 14-9）。这是因为氨基酸的体积较小，并且在某些情况下，它们非常相似。尽管挑战巨大，负载出错的概率却是非常小的；通常，1000 个 tRNA 分子中只有不到一个 tRNA 负载了错误的氨基酸。在某些情况下，如何实现这一高精确度是非常容易理解的。例如，Cys 和 Trp 在体积、构形和化学组成上都有明显的差异。即便是结构相似的 Tyr 和 Phe（图 14-9a），由于 Tyr tRNA 合成酶与前者的羟基而非后者的羟基能形成较强的和能量上有利的氢键，使得其能够有效地鉴别排斥 Phe。

更大的挑战在 Ile 和 Val 之间，它们仅有一个亚甲基（methylene）基团的差异（图 14-9b）。缬氨酰 tRNA 合成酶可通过其催化口袋的空间位阻作用排斥 Ile，因为 Ile 的体积大于 Val。相反，Val 应该能轻易地滑入异亮氨酰 tRNA 合成酶的催化口袋中。虽

然这两种氨基酸都能结合在异亮氨酰 tRNA 合成酶的氨基酸结合位点上，（异亮氨酰 tRNA 合成酶）与 Ile 额外的亚甲基基团的作用能够产生额外的-2 到-3kcal/mol 的自由能（表 3-1）。如第 3 章所述，如果这两种氨基酸的浓度相同，即使相对微小的自由能差异也能使（异亮氨酰 tRNA 合成酶）与 Ile 的结合约 100 倍可能于（酶）与 Val 的结合。因此，Val 大约可与 1%的异亮氨酰 tRNA 合成酶结合。但是，这是一个不能接受的高错配率。我们已经说过，实际上的错配率小于 0.1%。这样高的保真度是如何实现的呢?

某些氨基酰 tRNA 合成酶利用编辑口袋来高精度地负载 tRNA

提高氨基酰 tRNA 合成酶保真度的一个普遍的机制是对负载反应的产物进行校对，正如第 8 章中讨论过的 DNA 聚合酶的校对作用一样。例如，除了催化口袋（负责腺苷酰基化），异亮氨酰 tRNA 合成酶在催化口袋附近还有一个编辑口袋（酶上的深裂缝），用以对腺苷酰基化产物进行校对。AMP-Val（同样还有其他小氨基酸，如 Ala 等的腺苷酰基化物）能够进入这样的编辑口袋，在那里被水解成自由的 Val 和 AMP 而释放出来。相反，AMP-Ile 因太大而无法进入编辑口袋，因而不能被水解。其结果为，异亮氨酰 tRNA 合成酶能够对 Val 进行两次筛选：在最初的氨基酸结合和腺苷酰基化过程中（筛选因子约为 100），以及随后的腺苷酰基化氨基酸的编辑过程中（同样筛选因子约为 100），两者总共可进行大约 10 000 倍的筛选（即错配率约为 0.01% ）。

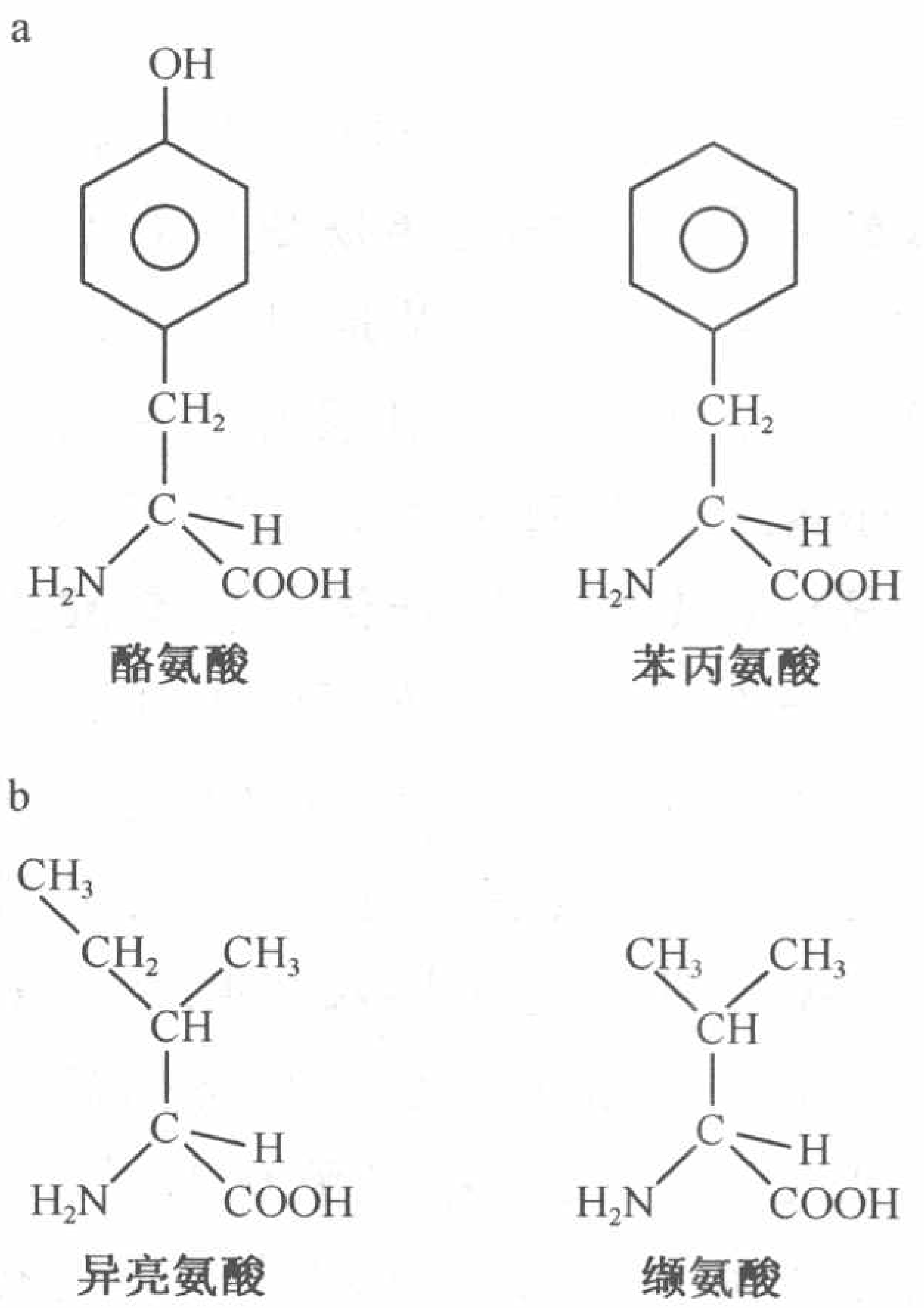


图 14-9　相似的氨基酸，不同的特征。

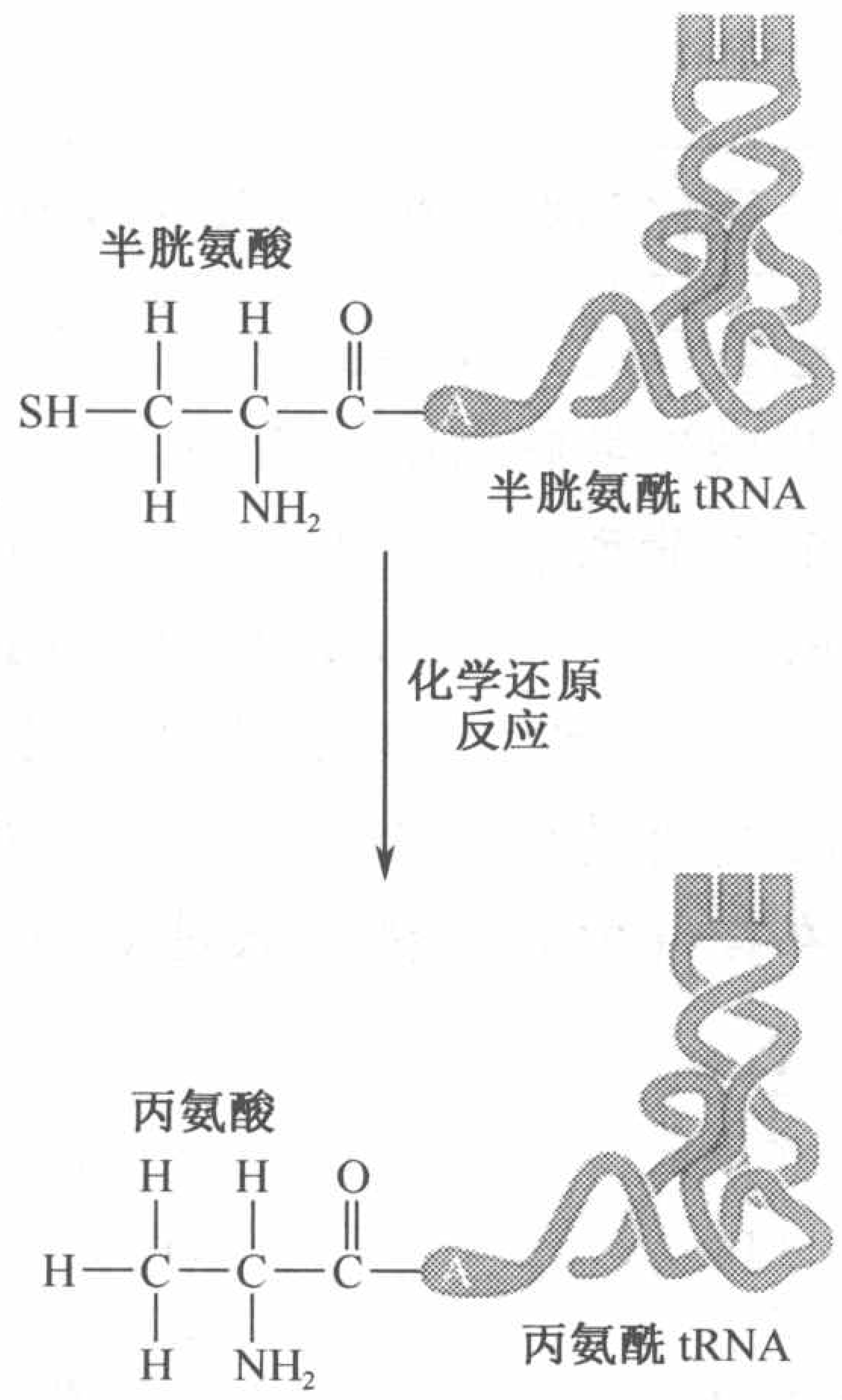


图 14-10　负载了 Cys 或 Ala 的半胱氨酰 tRNA。连接在半胱氨酰 tRNA 上的半胱氨酸的化学还原反应。

核糖体不能辨别 tRNA 的负载是否正确

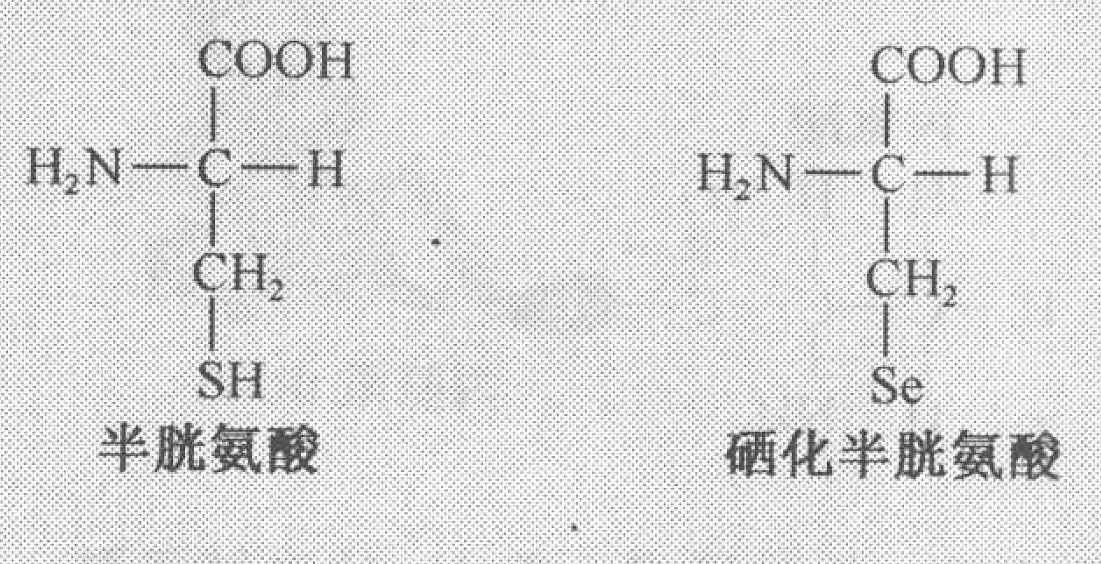
氨基酰 tRNA 合成酶承担如此多的责任以保证正确的氨基酸和正确的 tRNA 耦连的原因在于，核糖体不能辨别 tRNA 的负载是否正确。换句话说，核糖体只是“盲目地”接受任何表现出正确的密码子-反密码子相互作用的负载 tRNA，不管 tRNA 携带的是否是正确的氨基酸。

这个结论由两类实验支持：一类为遗传学的，另一类是生物化学的。遗传实验分离出一种反密码子上携带一个核苷酸替换的突变 tRNA。回想 tRNA 合成酶经常不依赖与反密码子的相互作用来识别相关的 tRNA 分子。这样，有些 tRNA 可以在反密码子上存在突变但仍然可以负载通常的相关氨基酸。但是，作为反密码子突变的后果，突变的 tRNA 将氨基酸带到错误的密码子上。换句话说，核糖体和其他与核糖体联合作用的辅助蛋白(将要简短地讨论)主要核对的是负载 tRNA 是否与 mRNA 具有正确的密码子-反密码子作用。核糖体和这些蛋白质极少阻止错误负载的 tRNA 将错误的氨基酸加到多肽链上。

一个经典的生化实验很好地说明了核糖体只识别 tRNA 而非其携带的氨基酸这个观点。看看负载半胱氨酸的半胱氨酰 $tRNA^{Cys}$（记住前缀指氨基酸，上标指 tRNA 的本质）。半胱氨酰 $tRNA^{Cys}$负载的 Cys 可通过化学方法变为 Ala 从而形成丙胺酰 $tRNA^{Cys}$（图 14-10）。当加入到不依赖于细胞的蛋白质合成体系中，丙胺酰 $tRNA^{Cys}$将 Ala 带到指定 Cys 插入的密码子上。因此，翻译机器主要依靠氨基酰 tRNA 合成酶的高保真度来保证每个 mRNA 正确解码（框 14-2）。

框 14-2 硒化半胱氨酸（selenocysteine）

某些蛋白质，如谷胱甘肽过氧（化）物酶和甲酸盐脱氢酶，含有一个独特的氨基酸，叫做硒化半胱氨酸（Sec），它是酶的催化中心的一部分。硒化半胱氨酸在 Cys 的硫原子处是微量元素硒（框 14-2 图 1）。有趣的是，硒化半胱氨酸不是通过翻译后进行化学修饰而掺入的（其他某些独特的氨基酸是通过这一方法掺入的如在胶原质中发现的羟基脯氨酸）。事实上，硒化半胱氨酸是从半胱氨酰-tRNA 合成酶负载的特殊 tRNA 上携带的 Cys 通过酶促反应而产生的。这个改变后的 tRNA 被用来把硒化半胱氨酸直接掺入到正在合成中的谷胱甘肽过氧（化）物酶中去。一个专用（类似于 EF-Tu；见下）翻译延伸因子把硒化半胱氨酰 tRNA 送达核糖体的（UGA）密码子，该密码子通常被识别为终止子。在 UGA 密码子处的掺入硒化半胱氨酸要求在 mRNA 的其他地方存在一个特殊的序列元素。因此，硒化半胱氨酸可以被认为是第 21 个氨基酸，它通过对细胞的标准翻译机器进行修饰而掺入到蛋白质中。



框 14-2 图 1 Cys 和 Sec 的结构。

核糖体

核糖体是指导蛋白质合成的大分子机器。与将核苷酸密码翻译成氨基酸密码所面临的额外挑战相一致，核糖体比合成 DNA 或 RNA 所需的最小的机器要大而复杂。实际上，单一的多肽可以执行 DNA 或 RNA 的合成（虽然 DNA 的复制和转录往往由更大的多亚基复合物来介导）。相比之下，聚合氨基酸的机器则至少由 3 种 RNA 分子和多于 50 种的不同蛋白质组成，其总分子质量大于 2500kDa。与 DNA 复制的速度相比较（每秒 200～1000 个碱基），翻译只能以每秒 2～20 个氨基酸的速度进行。

原核细胞中，转录和翻译的系统位于同一区域中。因此，核糖体能够在 mRNA 刚从 RNA 聚合酶中暴露出来时着手翻译。这样核糖体可以随着 RNA 聚合酶沿着转录体进行翻译（图 14-11）。因为 RNA 的 5′端首先合成，因此从 mRNA5′端开始翻译的核糖体可以在初始转录体刚刚在 RNA 聚合酶中暴露出来的时刻开始翻译。有趣的是，原核细胞中存在几个利用转录和翻译耦联现象进行基因表达调节的例子，我们将在第 18 章中对此进行讨论。

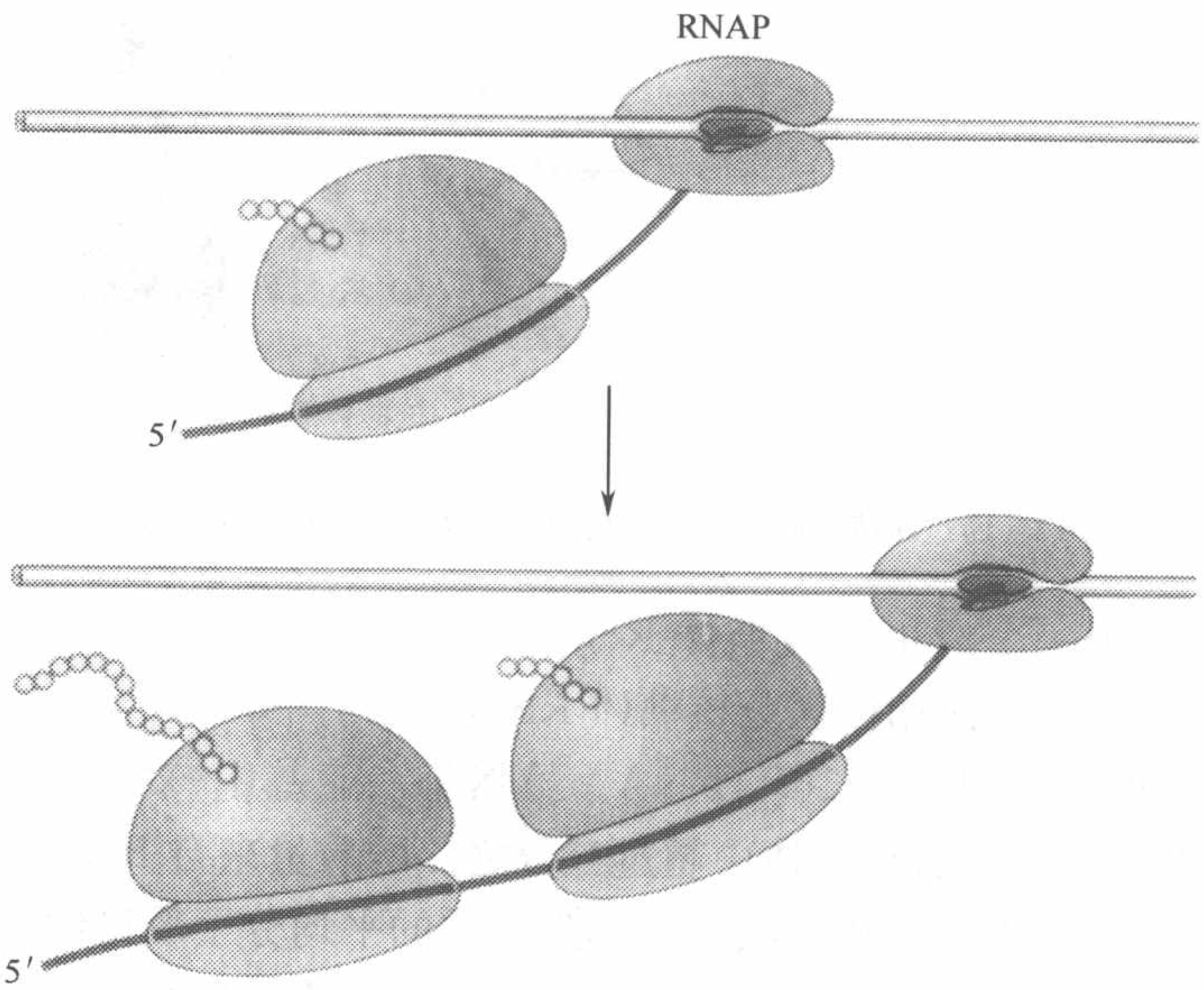


图 14-11　原核细胞的 RNA 聚合酶和核糖体作用于同一 mRNA。

虽然原核细胞中，（翻译的速度）相对于 DNA 合成的速度要慢，但核糖体仍可以赶上转录的速度。原核细胞典型的每秒 20 个氨基酸的翻译速度与 mRNA 的每秒 60 个核苷酸（20 个密码子）的翻译速度相当。这与 RNA 聚合酶每秒合成 50～100 个核苷酸的速度也相仿。

与原核细胞的情况不同，真核细胞的翻译完全与转录分开。这些事件在细胞的不同

区域里进行：转录发生在细胞核，而翻译发生在细胞质中。也许由于不与转录耦联，真核细胞的翻译以每秒 2～4 个氨基酸的较慢的速度进行。

核糖体是由一个大亚基和一个小亚基组成的

核糖体是由 RNA 和蛋白质组成的大亚基和小亚基两个部件组成的。大亚基含有肽基转移酶中心（peptidyl transferase center），负责肽键的形成。小亚基含有解码中心（decoding center），负载氨基酸的 tRNA 在此阅读或"解码"mRNA 的密码子单元。

按照惯例，大、小亚基的命名是根据离心时的沉降速率而定的（图 14-12）。测量沉降速率的单位是 Svedberg（S；S 值越大，沉降速率越快分子越大），这是以超速离心的发明者 Theodor Svedberg 命名的。原核细胞中，大亚基的沉降速率是 50 Svedberg，因此相应地称为 50S 亚基，而小亚基称为 30S 亚基。完整的原核细胞的核糖体称为 70S 核糖体。注意 70S 小于 50S 和 30S 相加的和！对这一明显矛盾现象的解释是沉降速率是由形状和大小共同决定的，因而不是绝对质量的精确测量。真核细胞的核糖体稍微大些，由 60S 和 40S 两个亚基组成，合在一起形成 80S 的核糖体。

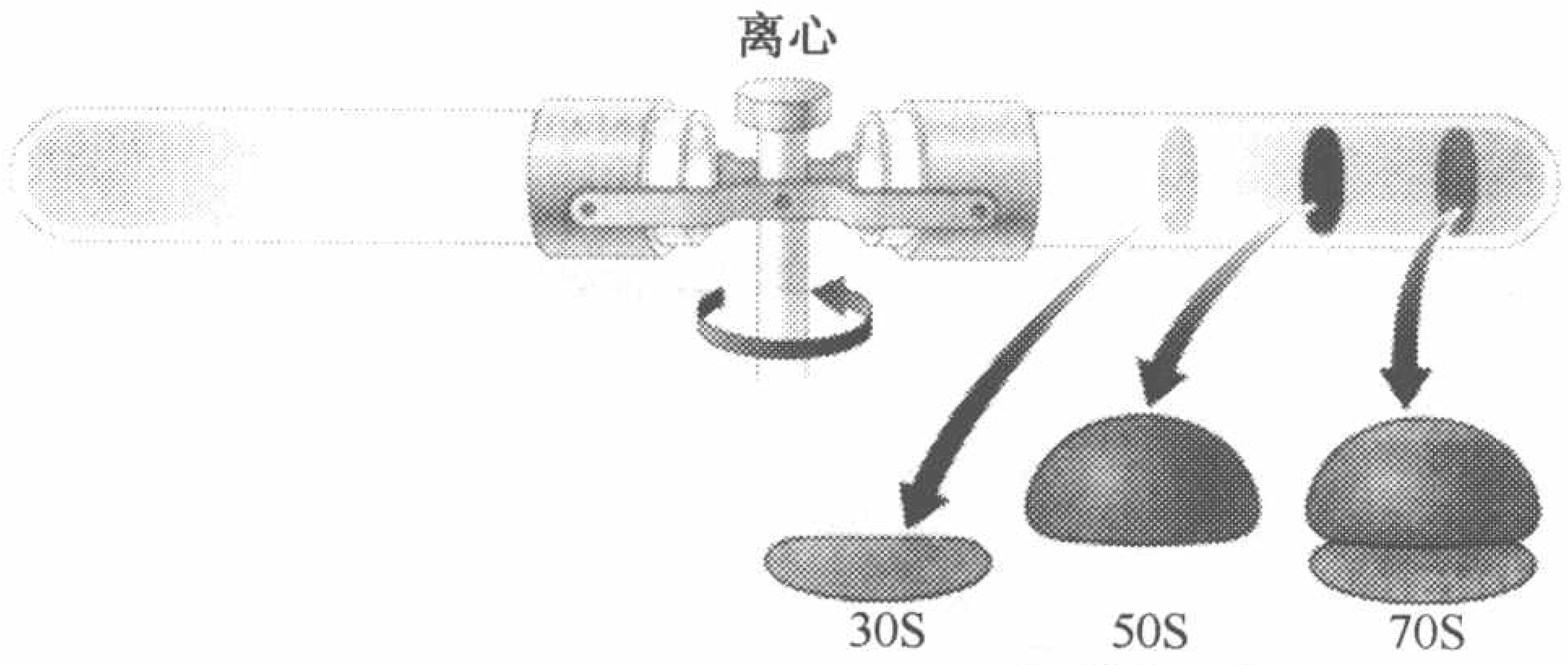


图 14-12 通过超速离心沉淀从完整的核糖体中分离核糖体亚基。

大、小亚基都由一个或更多核糖体 RNA（rRNA）和许多核糖体蛋白质组成（图 14-13）。Svedberg 单位再次用以区分核糖体 RNA。细菌中，50S 亚基含有 5S rRNA 和 23S rRNA，而 30S 亚基只含有 16S rRNA。虽然每一亚基中核糖体蛋白质要远远多于核糖体 RNA，但是原核细胞核糖体中超过三分之二的质量来自 RNA。这是真的，因为核糖体蛋白质比较小（细菌小亚基的核糖体蛋白质的平均分子质量约为 15kDa）。相反，16S 和 23S rRNA 比较大。我们想想，平均来说，一个核苷酸的分子质量是 330Da。因此，由 2900 个核苷酸组成的 23S rRNA 的分子质量即接近 1000kDa。

大、小亚基在翻译的每次循环过程中都经历结合与分离

蛋白合成一次，翻译机器就经历一系列特异的事件：核糖体的大、小亚基相互之间及与 mRNA 相结合，翻译 mRNA，完成蛋白质合成后相互分离。这种结合和分离相间的顺序称为核糖体循环（ribosome cycle，图 14-14）。简而言之，当 mRNA 和起始 tRNA 结合于游离的小亚基上，翻译过程就开始了。随后这"小亚基-mRNA-起始 tRNA"

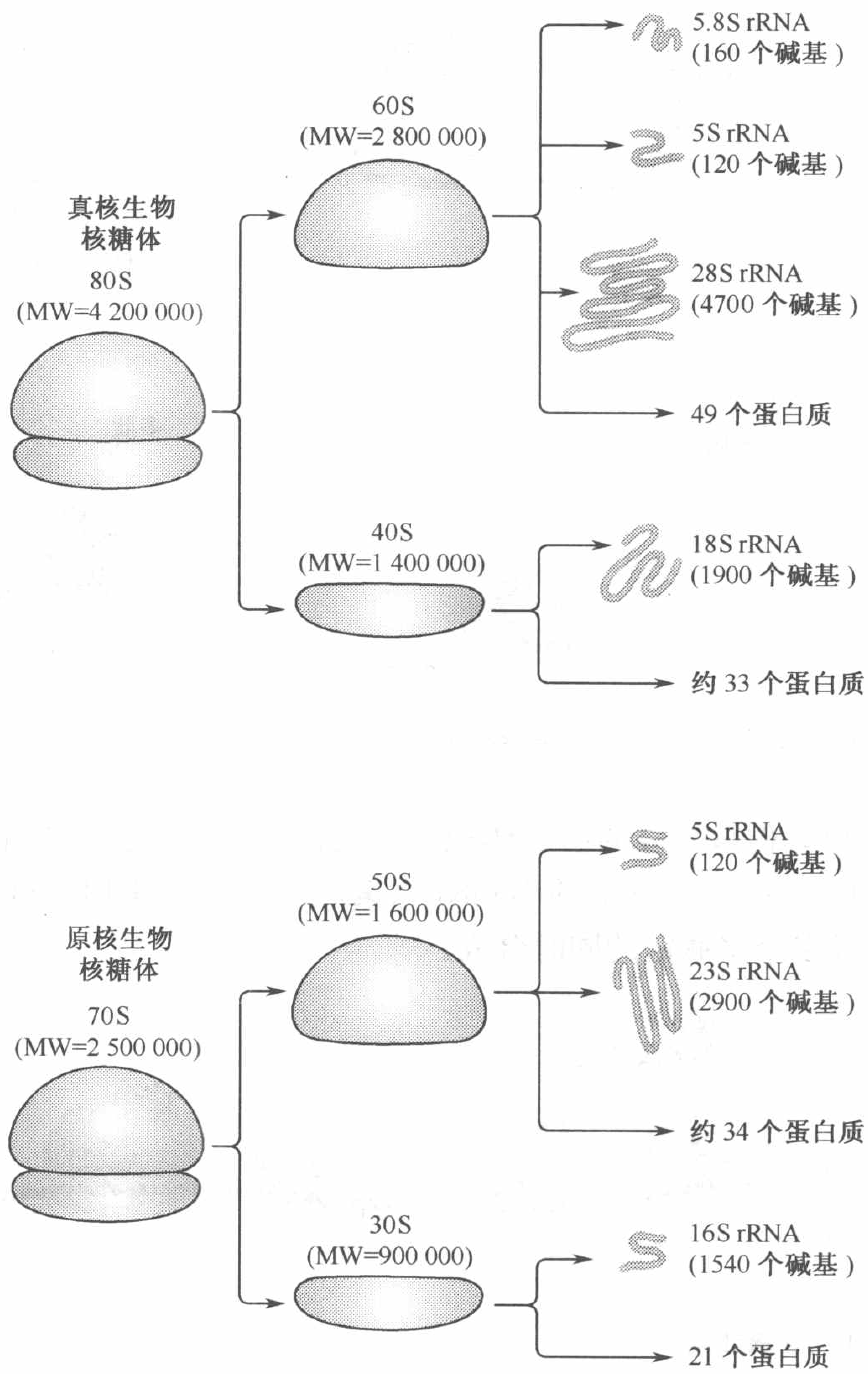


图 14-13 原核细胞和真核细胞的核糖体组成。标明了不同亚基的 rRNA 和蛋白质组分。还标明了各亚基的 rRNA 大小和蛋白质数目。

的复合体吸引大亚基结合，从而形成完整的、大小亚基间夹着 mRNA 的核糖体。下一步蛋白质合成启动，始于 mRNA 5′端的起始密码子并向 mRNA3′端移动。随着核糖体从一个密码子易位到另一个密码子，一个接一个的负载 tRNA 进入核糖体的解码和肽基转移酶中心。当移动的核糖体碰到终止密码子时，合成完成的多肽链就被释放出来，核糖体大、小亚基分离，各自离开 mRNA。这些分离的亚基就可以结合到新的 mRNA 分子上，重复蛋白质合成的循环。

虽然一个核糖体一次只能合成一条多肽，但每个 mRNA 分子却能同时被多个核糖体同时翻译（简便起见，我们假设 mRNA 是单顺反子的）。结合多个核糖体的 mRNA 称为多核糖体（polyribosome 或 polysome，图 14-15）。一个核糖体大约与 mRNA 上的

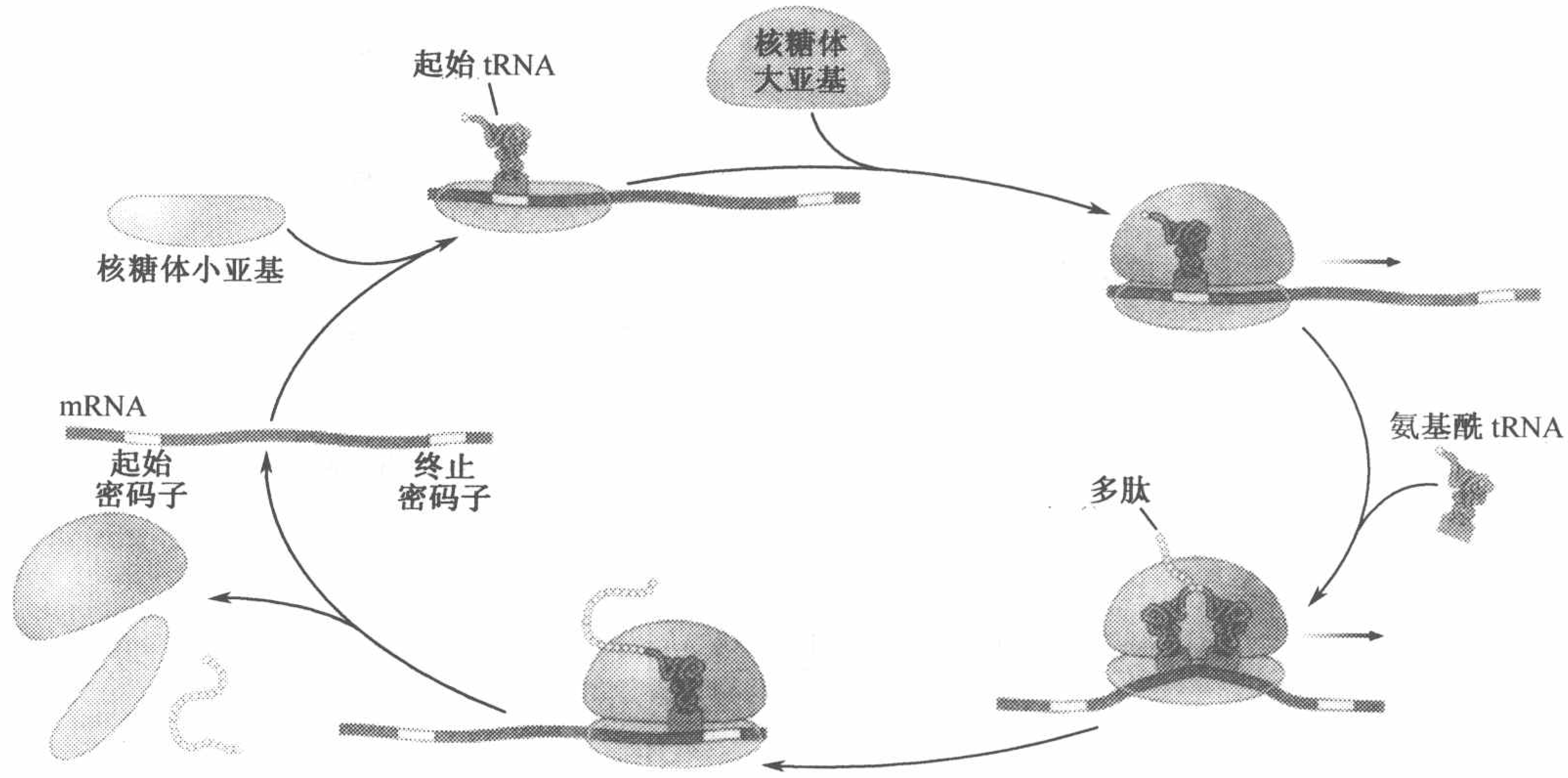


图 14-14 翻译的总体过程：核糖体循环。

30 个核苷酸相接触，但核糖体的巨大体积只能允许 mRNA 上每 80 个核苷酸结合一个核糖体。即便这样，一个 1000bp 的小 ORF（编码 35kDa 的蛋白）可以结合多于 10 个核糖体，从而指导多个多肽链的同时合成。

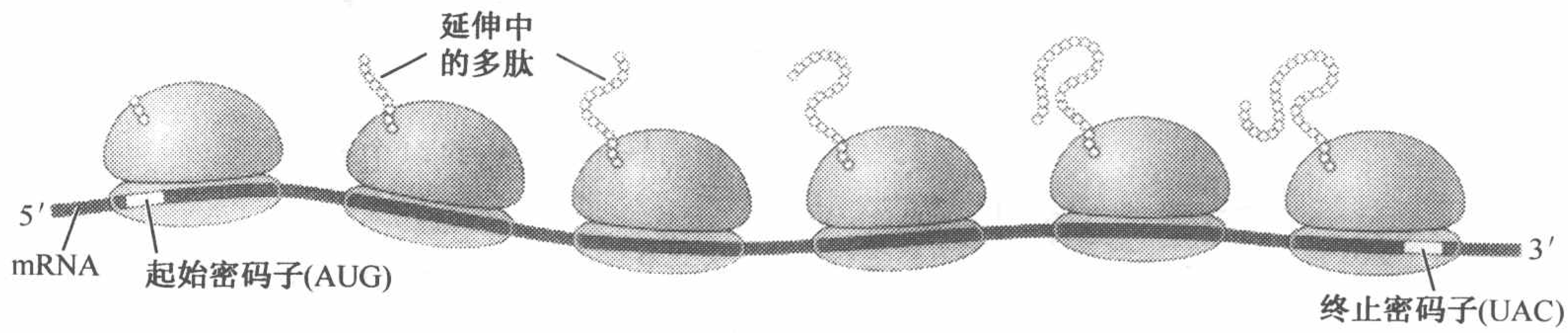


图 14-15 多核糖体。

多个核糖体作用于单一 mRNA 分子的能力解释了细胞中 mRNA 为什么相对较少（一般仅占总 RNA 的 1%～5%）。如果一个 mRNA 一次只能被一个核糖体翻译，那么，一个典型细胞中只有约 10%的核糖体能参加蛋白质的合成。但是多个核糖体和每个 mRNA 分子的结合确保在任何时候大多数的核糖体都参与了翻译。

新的氨基酸连接在延伸肽链的 C 端

我们已经知道，多核苷酸链和多肽链都有内在的极性。因此，我们可以问这些链的哪一端首先合成。从第 8 章和第 12 章中我们知道 DNA 和 RNA 都是将新的核苷酸加到延伸多核苷酸链的 3′端（通常称为 5′→3′方向合成）。

多肽链的合成方向如何呢？这最初由 Dintzis 所做的一个经典的实验所确定，该实验已在第 2 章有所描述。这一实验发现每一个新的氨基酸必须加在延伸多肽链的 C 端

(通常称为 N→C 方向合成)。如下一小节将要描述的，这样的合成方向由蛋白质合成的化学特性直接决定的。

肽键是通过将正在延伸的多肽链从一个 tRNA 转移到另一个 tRNA 而形成的

核糖体催化了单一的化学反应：肽键的形成。这一反应发生在延伸多肽链 C 端的氨基酸残基和将要加入的氨基酸之间。多肽链和将要加入的氨基酸都结合在 tRNA 上，其结果是，在肽键的合成过程中，不断增长的多肽链持续地结合在 tRNA 分子上。

每轮氨基酸加入的实际底物是两种负载的 tRNA：氨基酰 tRNA 和肽基酰 tRNA (peptidyl-tRNA)。如本章前面所讨论的 (见氨基酸与 tRNA 的结合章节)，氨基酰 tRNA 结合在氨基酸 3′端的羧基上。肽基酰 tRNA 也以同样的方式 (3′端) 结合在多肽链的 C 端上。氨基酰 tRNA 与氨基酸之间的键在下一个肽键的形成过程中是不断裂的。相反，当延长的多肽链在与负载氨基酰 tRNA 上的氨基酸的氨基进行结合而形成一个新的肽键时，肽基酰 tRNA 与多肽链之间的键断裂了。

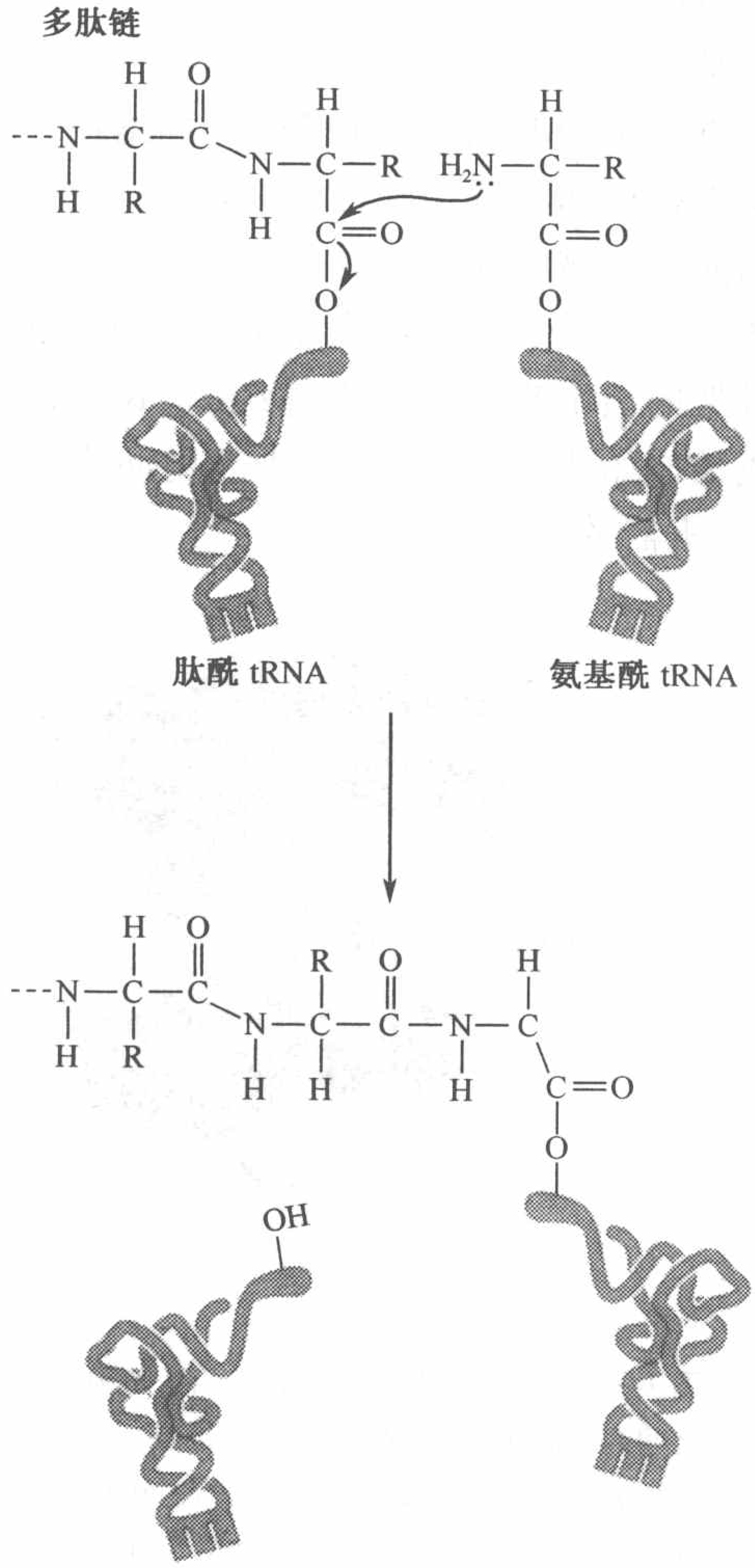


图 14-16　肽酰转移酶反应。

为了催化肽键的形成，这两个氨基酸的 3′端被带到核糖体上相邻的位置。这一位置效应可以使氨基酰 tRNA 的氨基基团攻击肽基酰 tRNA 上多肽 C 端的羧基集团，形成一个新的肽键 (图 14-16)。多肽合成的这一方式产生了两个后果。第一，这个肽键形成的原理要求蛋白质的 N 端先于 C 端合成；第二，多肽链从肽基酰 tRNA 转移到氨基酰 tRNA 上。因此，新肽键形成的反应称为肽转移酶反应 (peptidyl transferase reaction)。

有趣的是，肽键的形成并没有伴随三磷酸核苷的水解。这是因为肽键的形成是由连接多肽链和 tRNA 的高能酰键的断裂来驱动的。请回忆一下，这一高能酰键是在 tRNA 负载氨基酸时由 tRNA 合成酶催化产生的。tRNA 负载氨基酸的反应伴有一个 ATP 分子的水解。这样，肽键形成所需的能量来自于 tRNA 负载氨基酸时所水解的那一个 ATP 分子 (图 14-16)。

核糖体 RNA 是核糖体中的结构和催化决定因子

虽然核糖体和它的基本功能早在 40 年以前就发现了，但近年来对原核细胞核糖体的高精度的三维结构的确定大大加快了我们对这一分子机器工作机制的认识（图 14-17）。也许其中最重要的结果就是发现核糖体 RNA 并不单单是核糖体的结构成分，甚而，它们还直接负责核糖体的关键功能。最明显的证据就是证实肽基转移酶中心完全是由 RNA 组成的，这将在下面详细讨论。RNA 也在核糖体小亚基功能中扮演中心的角色。负载 tRNA 的反密码子环和 mRNA 的密码子都是与 16S rRNA 相作用，而不是与小亚基的核糖体蛋白质相作用。

进一步说明 RNA 在核糖体的结构和功能的重要性的证据是，大多数的核糖体蛋白质位于核糖体的周边，而非其内部（图 14-17 及网上结构教学材料 14-1）。核糖体的核心功能结构域（肽基转移酶中心及解码中心）完全或者绝大部分由 RNA 组成。某些核糖体蛋白质的一部分确实也伸入到亚基的核心，但它们的作用似乎是通过屏蔽其磷酸戊糖骨架的负电来稳定紧密缠绕的 rRNA。实际上，现代核糖体很可能是从完全由 RNA 组成的原始蛋白质合成机器进化而来，而核糖体蛋白质的加入是为了提升这种原始 RNA 机器的功能。

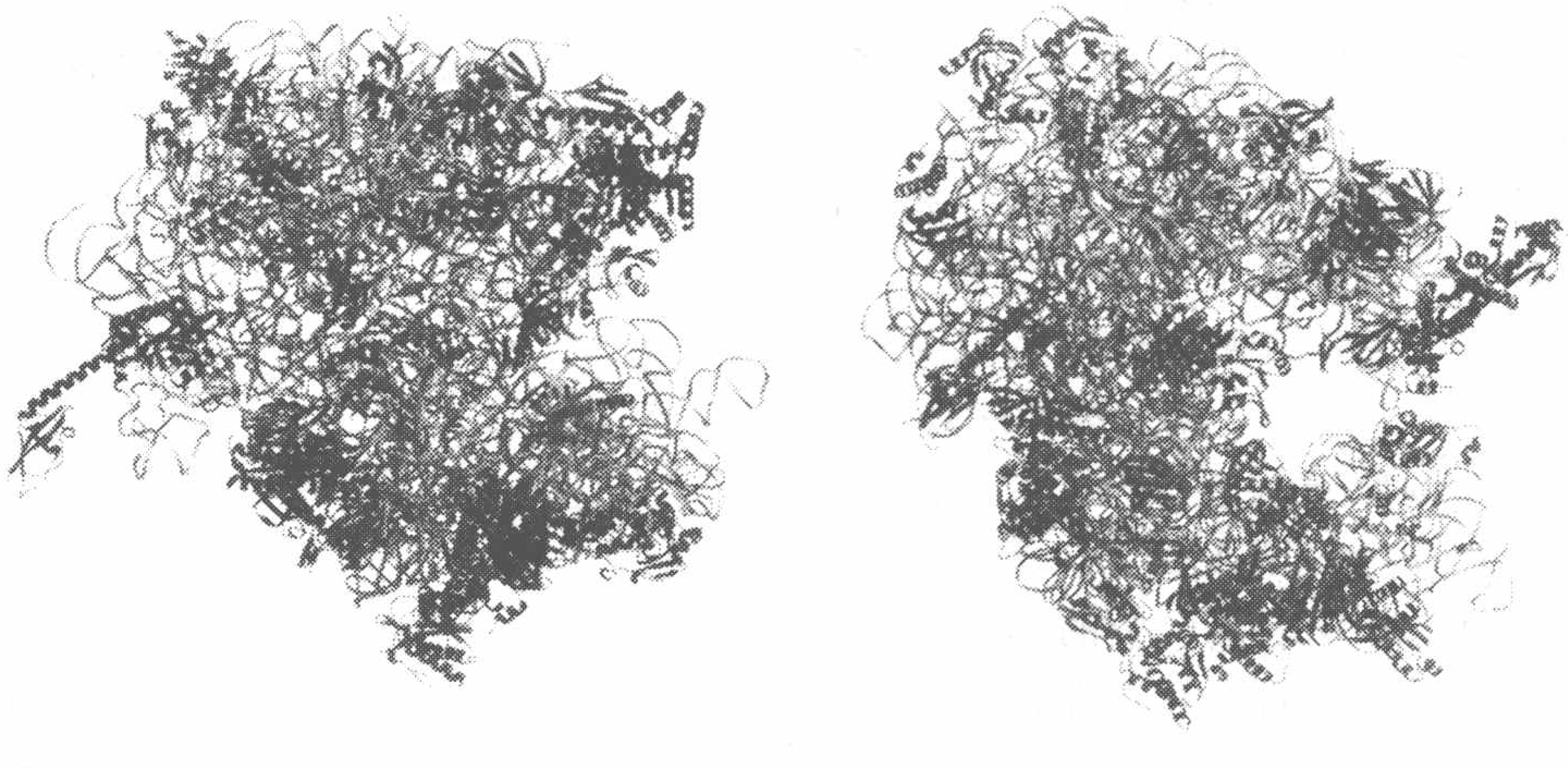


图 14-17 核糖体的两个视图。这两个图中，50S 亚基都位于 30S 亚基之上。右图中 50S 和 30S 亚基之间的空洞代表了 tRNA 结合的位点（图 14-19b）。50S 亚基的 RNA 部分以灰色表示而蛋白质以紫红色表示。30S 的 RNA 部分以淡蓝色表示而蛋白质以深蓝色表示。（Yusupov M. M. et al. 2001. Science 292：883-896.）图像使用 MolScript、BobScript 和 Raster 3D 软件制作。

核糖体有 3 个 tRNA 结合位点

为了进行肽基转移酶反应，核糖体必须能够同时结合至少两个 tRNA。实际上核糖体含有 3 个 tRNA 结合位点，分别为 A、P 和 E 位点（图 14-18 和图 14-19）。A 位点（A site）是氨基酰 tRNA 的结合位点，P 位点（P site）是肽基酰 tRNA 的结合位点，E 位点（E site）是延伸的多肽链转移到氨基酰-tRNA 后被释放的 tRNA 的结合位点（E 指 exit，退出）。

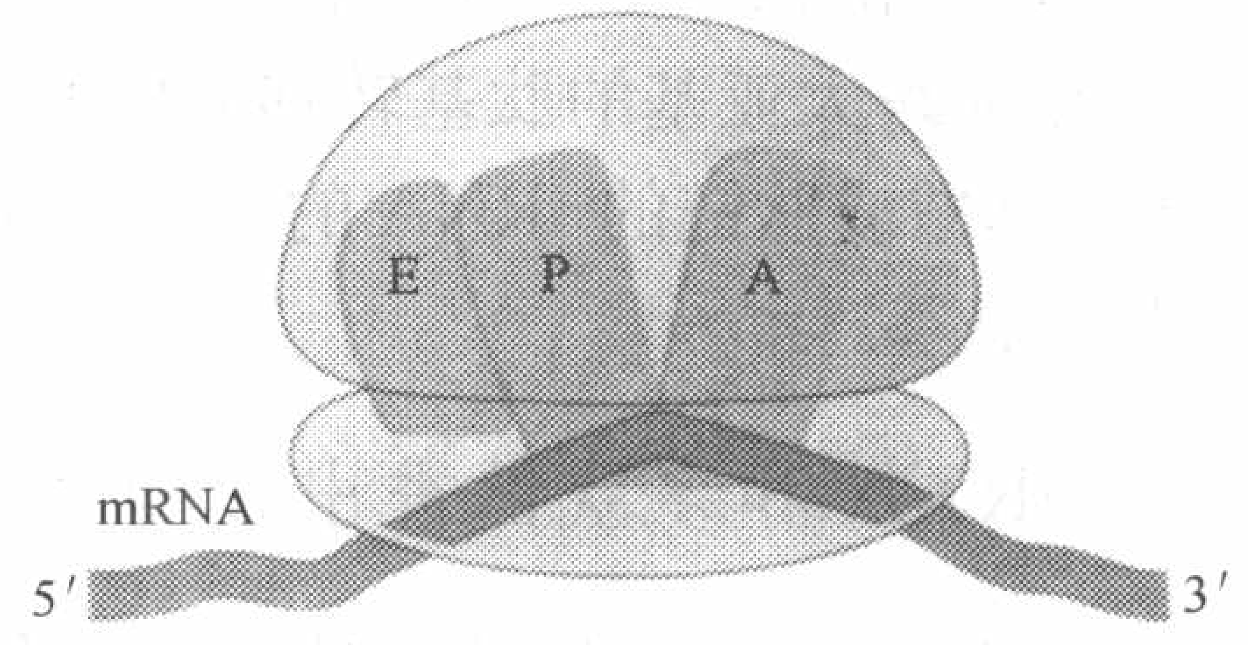


图 14-18 核糖体有 3 个 tRNA 结合位点。核糖体示意图显示横跨 2 个亚基的 3 个 tRNA 结合位点（E、P 和 A）。

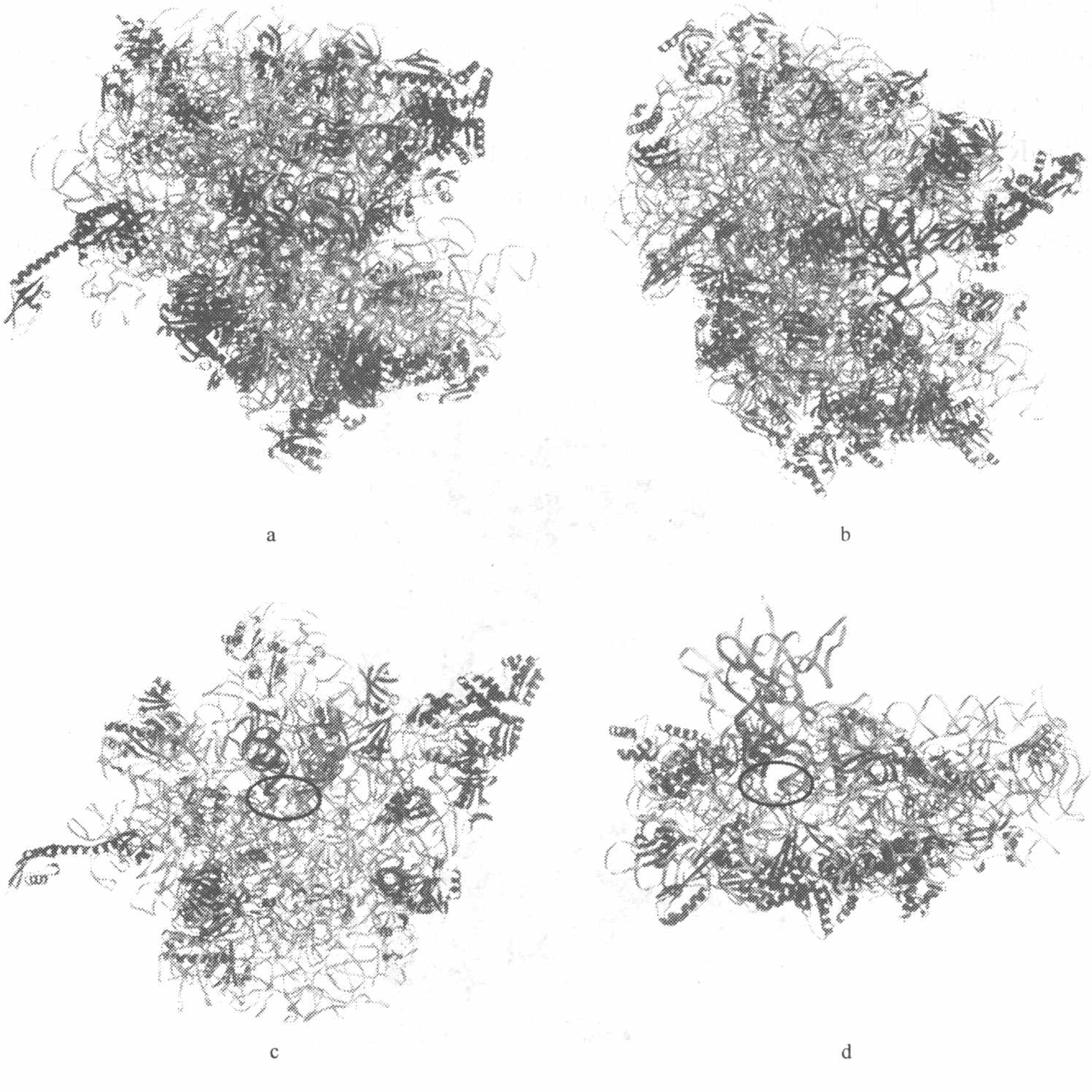


图 14-19 包含 3 个结合 tRNA 的核糖体三维结构示意图。E、P 和 A 位点的 tRNA 分别以黄、红和绿色表示。大、小亚基的 RNA 和蛋白质组分的颜色与图 14-17 相同。a 和 b 两图显示核糖体与 tRNA 结合于 E、P 和 A 位点，请注意 a 和 b 图的核糖体与图 14-17 的核糖体相对应。c. 分离的 50S 亚基结合于 tRNA，有如在整个核糖体中的样子（本图的角度是从小亚基向大亚基看）肽酰转移酶中心圈出来了。d. 分离的 30S 亚基结合于 tRNA 有如在整个核糖体中的样子，解码中心圈出来了。（Yusupov M. M. et al. 2001. Science 292：883-896.）图像使用 MolScript、BobScript 和 Raster 3D 软件制作。

每一个 tRNA 结合位点都在核糖体的大、小亚基的交界面形成（图 14-19a 和 b）。因此，结合核糖体的 tRNA 能够横跨大亚基的肽基转移酶中心（图 14-19c）和小亚基的解码中心（图 14-19d）。耦联氨基酸或肽链的 tRNA 的 3′端紧邻大亚基，结合核糖体的 tRNA 的反密码子环则紧接着小亚基。

穿过核糖体的通道可以使 mRNA 和多肽链进出核糖体

解码中心和肽基转移酶中心都埋于完整的核糖体内。但是，翻译过程中，mRNA 必须成串珠状通过解码中心，新合成的多肽链也必须从肽基转移酶中心离开。这些聚合体（如 mRNA）是如何进入和离开核糖体的呢？这个问题的答案就在核糖体的结构中，它展示进出核糖体的“通道”。

mRNA 通过小亚基的两个狭窄的通道进出解码中心。进入通道的宽度只能允许未配对的 RNA 通过。这个特征保证了 mRNA 以一种单链的形状进入解码中心，这是通过解除 mRNA 分子内可能已经形成的碱基配对相互作用而达到的。进出通道之间有一个tRNA可进入的区域，毗邻的密码子在此能够分别与 A 位点的氨基酰 tRNA 和 P 位点的肽基酰 tRNA 结合。有趣的是，mRNA 的两个密码子之间存在明显的扭结（kink），有利于维持正确的可读框（图 14-20）。这个扭结使核糖体易位后留在空着的 A 位点的

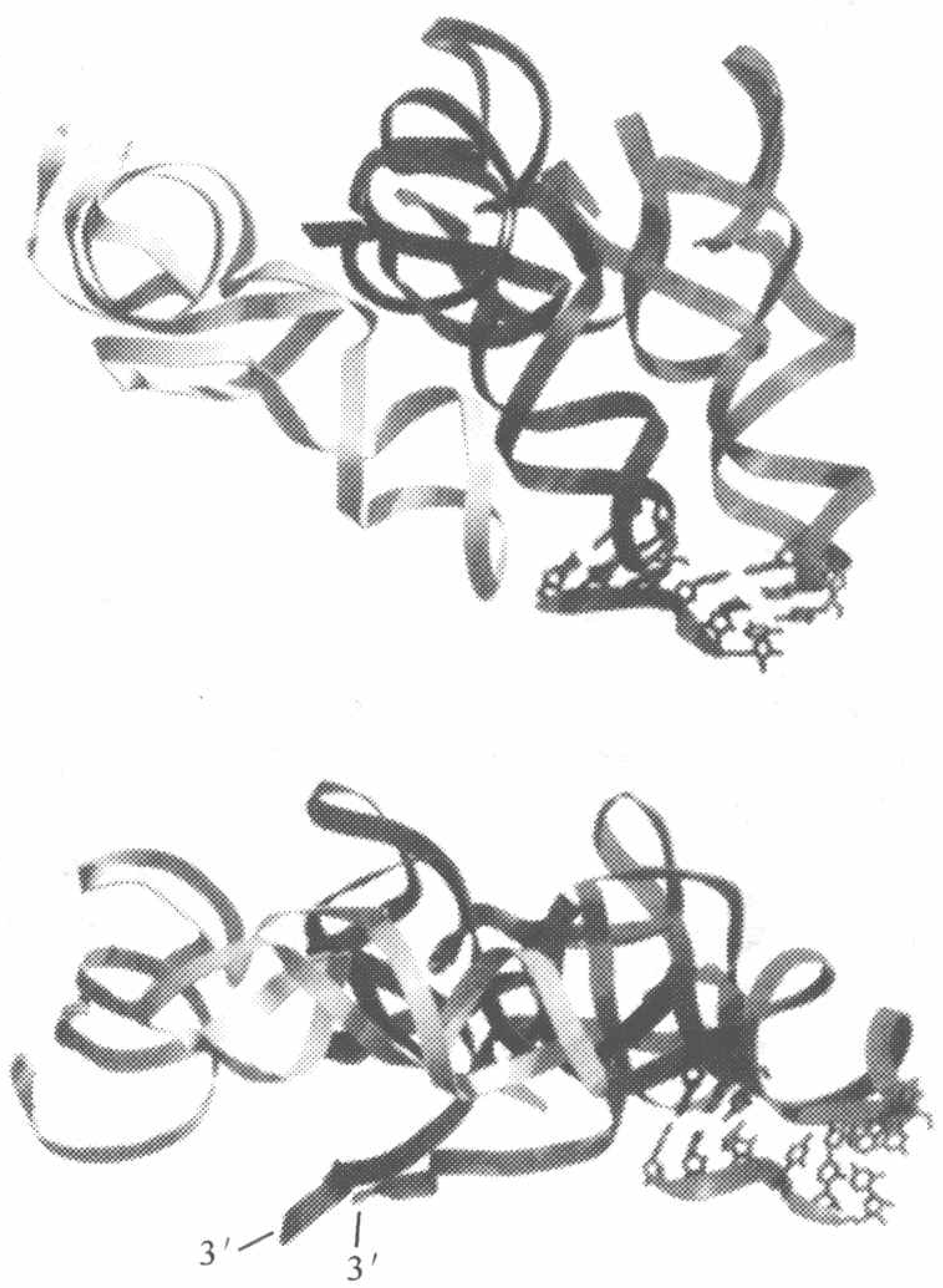


图 14-20 核糖体内 mRNA 与 A、P 位点的 tRNA 的相互作用。两图表示 mRNA 和 tRNA 在核糖体内的结构。为清晰起见，核糖体未示出。E、P 和 A 位点的 tRNA 分别以黄、红和绿色表示，mRNA 以蓝色表示。只有参与密码子-反密码子相互作用的碱基示出。mRNA 上明显的扭结清晰地区别 A 和 P 位点的密码子，在下图可看出 A 和 P 位点的 tRNA 的 3′端之间非常接近。（Yusupov M. M. et al. 2001. Science 292：883-896.）图像使用 MolScript、BobScript 和 Raster 3D 软件制作。

密码子处于一个独特的位置，阻止随后进入的氨基酰-tRNA 与毗邻该密码子的碱基相接触。

大亚基上的另一个通道提供了新合成的多肽链离开（核糖体）的途径（图 14-21）。与 mRNA 的通道一样，这个通道的大小限制了多肽链的折叠。在这种情况下，新合成的多肽链只能在通道内形成 α 螺旋，而其他的二级结构（如 β 折叠）和空间的相互作用只能在多肽链离开大亚基后形成。由于这一原因，新合成的多肽链最终的三维结构只能在其从核糖体释放后形成。

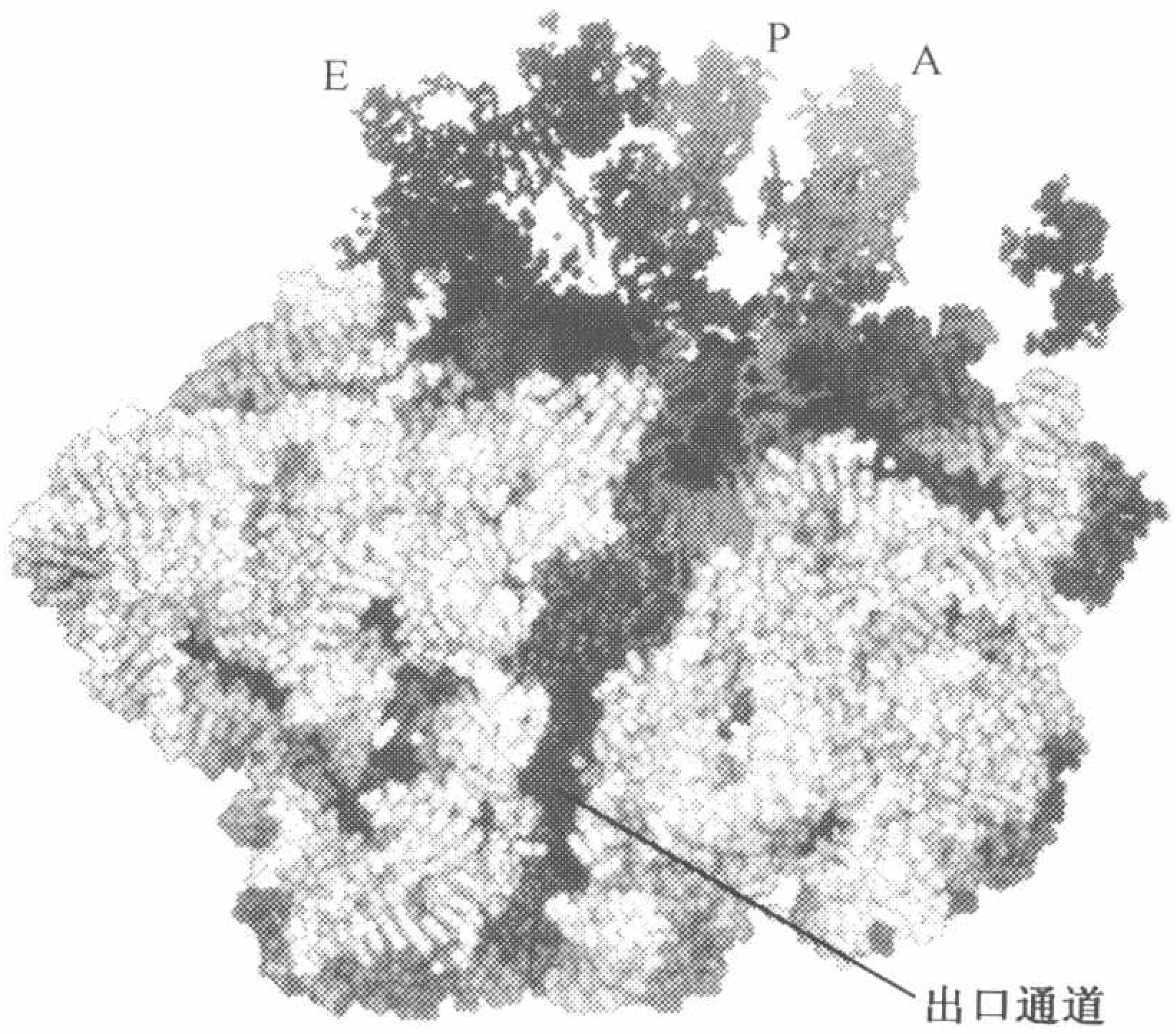


图 14-21 多肽出口通道。此图中，50S 亚基切成一半以显示多肽出口通道。rRNA 以白色显示而核糖体蛋白质以黄色显示。3 个结合的 tRNA 显示为：E 位点（褐色）、P 位点（紫红色）、A 位点（绿色）。紧接 A 位点 tRNA 的红色和金色的 rRNA 部分是肽酰转移酶中心的组成部分。（由 Courtesy of T. Martin Schmeing and Thomas Steitz 提供；采自 Schmeing T. M. et al. 2002. Nature Struct. Biol. 9：225-230.）

现在我们已经描述了翻译过程的 4 种主要组分，本章剩余的部分将集中描述翻译的每一个步骤。我们的描述将以翻译的 3 个步骤依次进行：多肽链合成的起始、多肽链的延伸和多肽链合成的终止。下面将要看到，原核生物和真核生物的蛋白质合成在策略上有重要的相同点和不同点。我们将在下面的各个小节中讨论这两种细胞的翻译机制的本质。正如 DNA 和 RNA 合成一样，虽然核糖体是蛋白质合成的中心，但各种辅助因子在翻译的各个阶段也起着重要的作用，而且它们也是蛋白质合成以迅速和正确的方式进行所必需的。

翻译的起始

翻译要成功地起始，必须有 3 个事件发生（图 14-22）。第一，核糖体必须被募集到 mRNA 上；第二，负载 tRNA 必须置于核糖体的 P 位点；第三，核糖体必须精确地定位在起始密码子上。其中第三条是关键的，因为这一步确定了 mRNA 翻译的可读框。

即使核糖体定位只发生一个碱基的易位也将导致一个完全不相关的多肽的合成（见上面和第 15 章有关信使 RNA 的讨论）。原核生物和真核生物 mRNA 结构的不同导致完成翻译起始方式的明显不同。我们先阐明原核细胞的起始情况，然后讨论真核细胞中所观察到的差异。

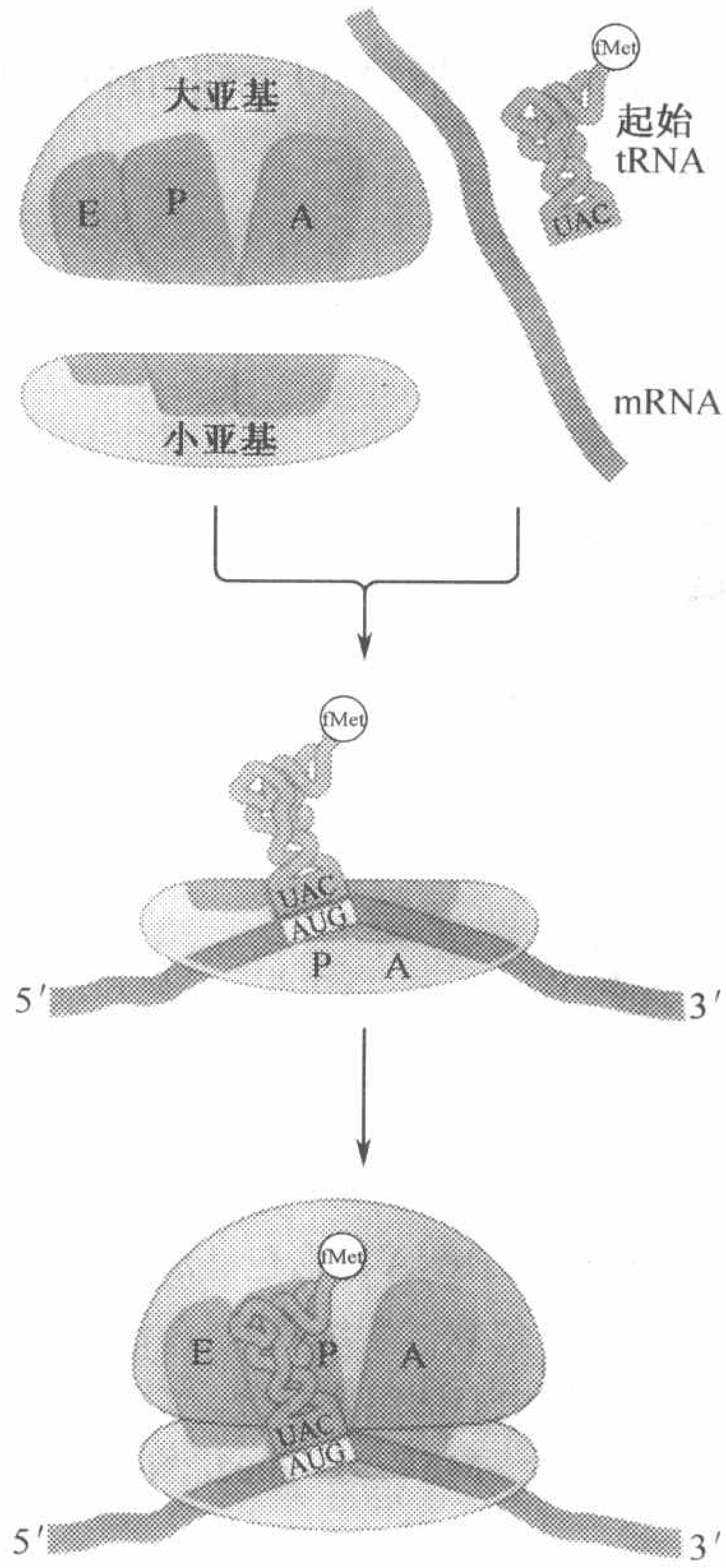


图 14-22 翻译起始事件的总体过程。

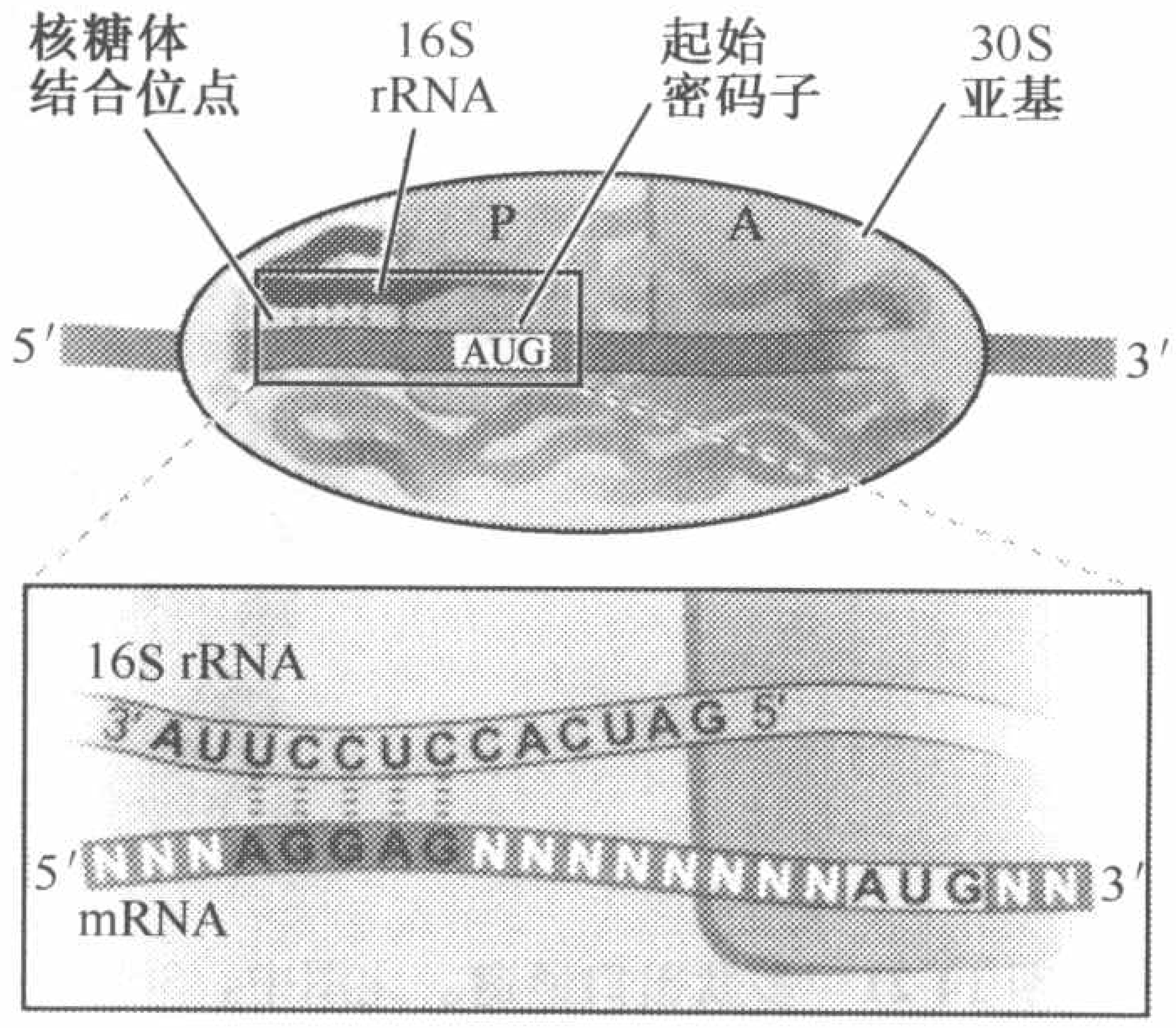


图 14-23 16S rRNA 作用于 RBS，将 AUG 定位于 P 位点。本图为 RBS 和起始 AUG 之间具有理想距离的 mRNA。这一距离将 AUG 置于 P 位点。许多 mRNA 没有这一理想距离，导致翻译效率的下降。其他一些 mRNA 完全缺失 RBS。

原核细胞的 mRNA 最初是通过与 rRNA 的碱基配对而募集到小亚基上

大、小亚基是一个一个地装配到 mRNA 上的。首先小亚基和 mRNA 结合。如在讨论 mRNA 结构时所述（图 14-2），对原核细胞来说，小亚基和 mRNA 的结合是由（mRNA 的）RBS 和（小亚基的）16S RNA 之间的碱基配对作用介导的（图 14-23）。对定位理想的 RBS 来说，小亚基处于 mRNA 上的恰当位置，使大亚基加入复合体时，起始密码子正好处于 P 位点。大亚基是在起始过程即将结束的时候加入复合体的，恰好在第一个肽键形成之前。因此，翻译起始的许多关键事件发生在完整的核糖体形成之前。

负载着修饰甲硫氨酸的特定的 tRNA 直接结合到原核细胞的小亚基上

通常，负载 tRNA 进入核糖体的 A 位点并在一轮肽键合成后才能到达 P 位点。但是，在起始阶段，负载 tRNA 直接进入到 P 位点。这一过程需要一个称为起始 tRNA（initiator tRNA）的特定 tRNA，它与起始密码子（通常是 AUG 或 GUG）碱基配对。

当 AUG 和 GUG 在可读框内出现时具有不同的含义，它们分别被 Met 的 tRNA（tRNA^{Met}）和 Val 的 tRNA（tRNA^{Val}）所识别和阅读（第 15 章）。起始 tRNA 上负载的既非 Met 亦非 Val。相反，它负载的是一种 Met 的修饰产物：在氨基基团上连着一个甲酸基（*N*-甲酰甲硫氨酸，*N*-formyl methionine）（图 14-24），负载此氨基酸的起始 tRNA 称为 fMet-tRNA$_i^{fMet}$。

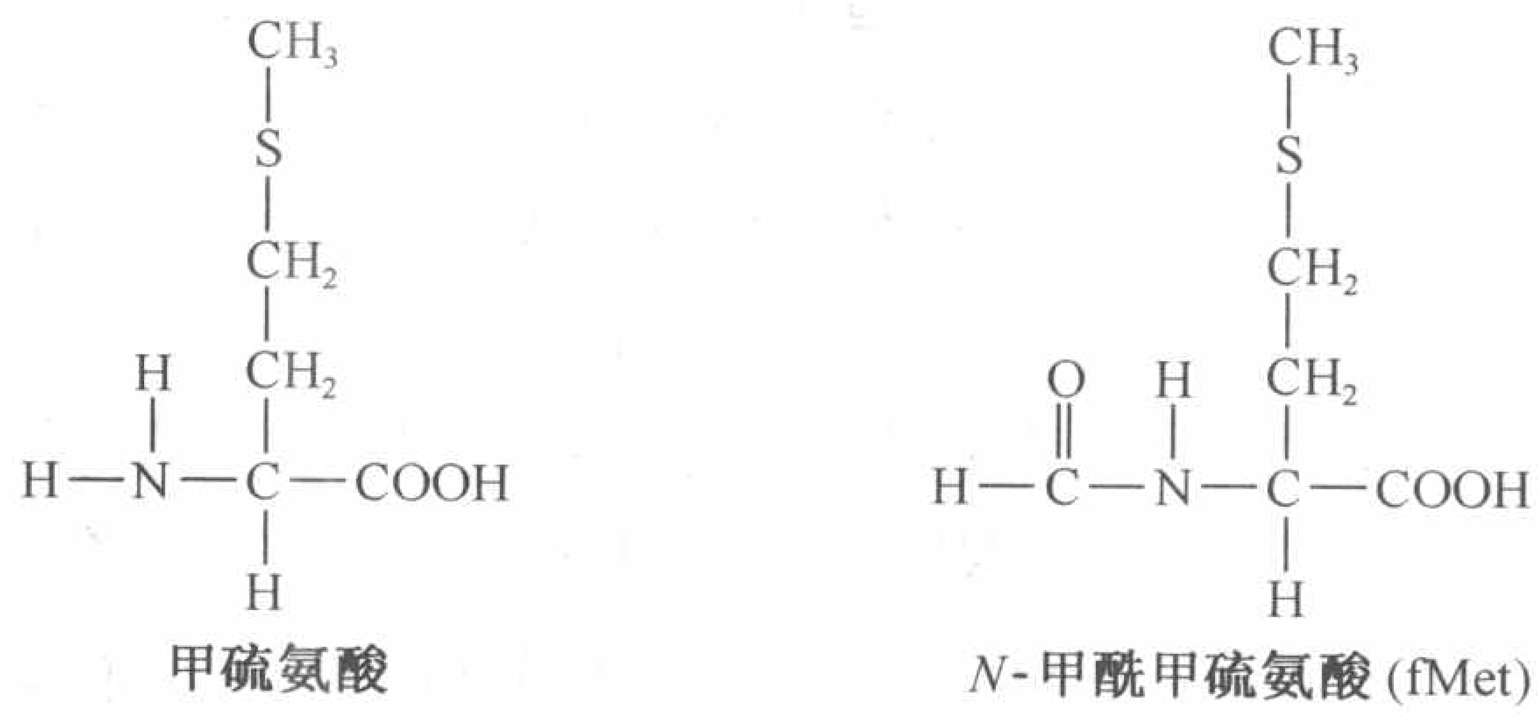


图 14-24 Met 和 *N*-fMet。

由于 *N*-fMet 是多肽链的第一个氨基酸，你也许认为所有原核生物的蛋白质在它们的氨基端都有甲酸基的结构。但事实并非如此，因为去甲酰化酶（deformylase）在多肽链合成的过程中或之后将这个甲酸基从氨基端去除掉。实际上，许多原核细胞的蛋白质甚至不是以 Met 开始的；氨肽酶（aminopeptidase）通常在氨基端切除 Met 以及另外一两个氨基酸。

三种起始因子指导包含 mRNA 和起始 tRNA 的起始复合物的装配

原核细胞的翻译始于小亚基，由 3 种翻译起始因子（translation initiation factor）——IF1、IF2 和 IF3 所催化，每一个因子催化起始过程的一个关键步骤：

- IF1 防止 tRNA 结合到小亚基中将形成 A 位点的位置上。
- IF2 是 GTP 酶（结合和水解 GTP 的蛋白质），它与起始过程的 3 个主要成分（小亚基、IF1 和 fMet-tRNA$_i^{fMet}$）相互作用。通过与这些成分相互作用，IF2 促使 fMet-tRNA$_i^{fMet}$和小亚基的结合，并阻止其他负载 tRNA 与小亚基结合。
- IF3 与结合小亚基并阻止其与大亚基重新结合。由于翻译起始需要游离的小亚基，因此 IF3 与小亚基的结合对翻译的新一轮的循环是十分重要的。IF3 在上一轮循环即将结束时开始与小亚基结合，并协助 70S 核糖体分解为大、小亚基。

每一个起始因子结合于小亚基的 3 个 tRNA 结合位点或其附近。与 IF1 防止负载 tRNA 结合到 A 位点的功能相一致，IF1 直接结合到小亚基中将形成 A 位点的位置上。IF2 结合 IF1 并跨过 A 位点到达 P 位点以结合 fMet-tRNA$_i^{fMet}$。最后，IF3 占据了将成为 E 位点的小亚基上的位置。这样，在小亚基的 3 个潜在的 tRNA 结合位点中，只有 P 位点在起始因子存在的条件下能够结合 tRNA。

随着 3 个起始因子的加入，小亚基已经准备好结合 mRNA 和起始 tRNA 了（图 14-25）。这两种 RNA 能以任何一种方向结合，并且互相独立。如上面所讨论的，（小亚基）结合 mRNA 需要(mRNA 的)RBS 和（小亚基的）16S RNA 之间的碱基配对。而

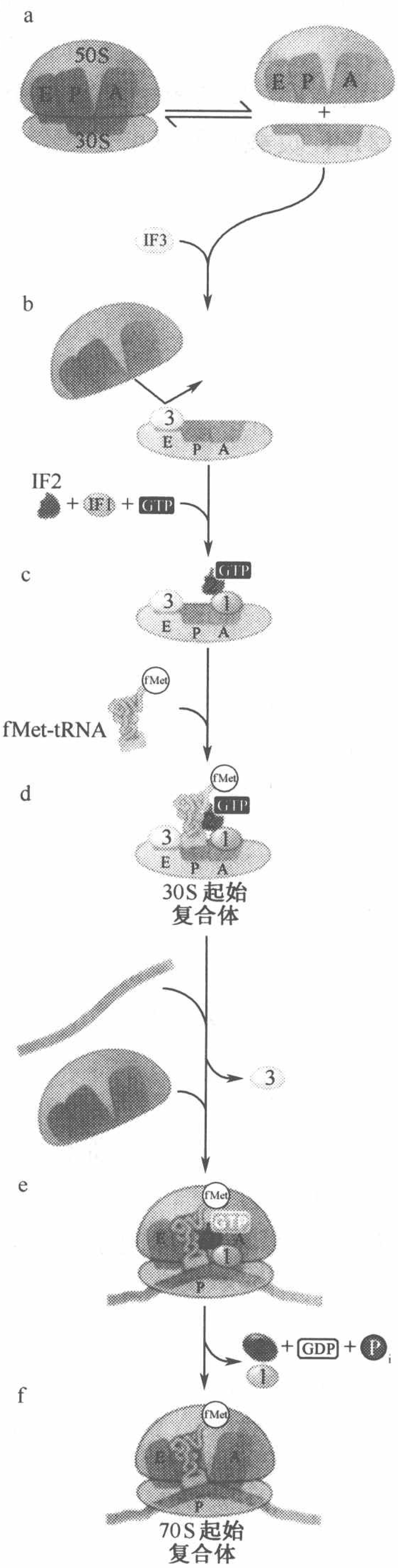


图 14-25 原核细胞的翻译起始过程。

$fMet\text{-}tRNA_i^{fMet}$结合小亚基是由结合 GTP 的 IF2 催化的，并且（一旦 mRNA 结合于小亚基上）由反密码子和 mRNA 的起始密码子之间的碱基配对作用促进。类似地，$fMet\text{-}tRNA_i fMet$ 和 mRNA 之间的碱基配对起到了定位起始密码子在 P 位点上的作用。

起始的最后步骤涉及大亚基的聚合以形成 70S 起始复合体（70S initiation complex）。当起始密码子和 $fMet\text{-}tRNA_i^{fMet}$的碱基配对后，小亚基的构象发生变化，导致 IF3 的释放。在 IF3 离开的情况下，大亚基可以自由地与小亚基及其负载的 IF1、IF2、mRNA 和 $fMet\text{-}tRNA_i^{fMet}$结合。与大亚基的结合激活了 IF2-GTP 的 GTP 酶活性，引起 GTP 的水解。水解后的 IF2-GDP 与核糖体和起始 tRNA 的亲和力降低，导致 IF2-GDP 和 IF1 从核糖体释放出来。这样，起始过程的最后产物是在 mRNA 的起始位点组装了一个完整的（70S）核糖体，其 P 位点有 $fMet\text{-}tRNA_i^{fMet}$，而 A 位点是空的。这个核糖体-mRNA 的复合体已准备好在 A 位点结合负载 tRNA，开始多肽链的合成。

真核细胞的 mRNA 通过 5′帽吸引核糖体

真核细胞翻译的起始在很多方面与原核细胞相同。它们都使用起始密码子和专门的起始 tRNA，都在大亚基的加入之前利用起始因子形成结合 mRNA的小亚基复合体。但是，真核细胞运用一个根本不同的方式来识别 mRNA 和起始密码子，这对真核细胞的翻译产生重要的结果。

在真核细胞中，小亚基在被募集到 mRNA 的 5′端帽子结构之前，已经与起始 tRNA 结合。然后以 5′→3′方向沿着 mRNA“扫描”直到遇见第一个 5′-AUG-3′（见前面章节中有关 mRNA 前导区 Kozak 序列的讨论），该密码子被识别为起始密码子。因此，在大多数情况下（框 14-3），只有第一个 AUG 才能作为真核细胞翻译的起点。请注意这一起始方式是和大多数真核mRNA编码一个多肽链（单顺反子，monocistronic）的事实相一致；对内部起始密码子的识别通常是不必要的，也是不可能的。

正如我们在其他分子过程（例如转录时对启动子的识别）中所看到的，真核细胞较之原核细胞需要更多的辅助蛋白来完成起始过程。整个起始过程可分为 4 步。首先，与在原核生物中不同，在真核细胞中，起始 tRNA 与核糖体小亚基的结合总是先于其与 mRNA 的结合（图 14-26a）。其次，另一组辅助因子介导对 mRNA 的识别。第三，结合于起始 tRNA 的核糖体在 mRNA 上“扫描”以搜寻第一个 AUG 序列。最后，核糖体的大亚基在起始 tRNA 与起始密码子碱基配对之后被招募进来。现在，我们详述这些事件。

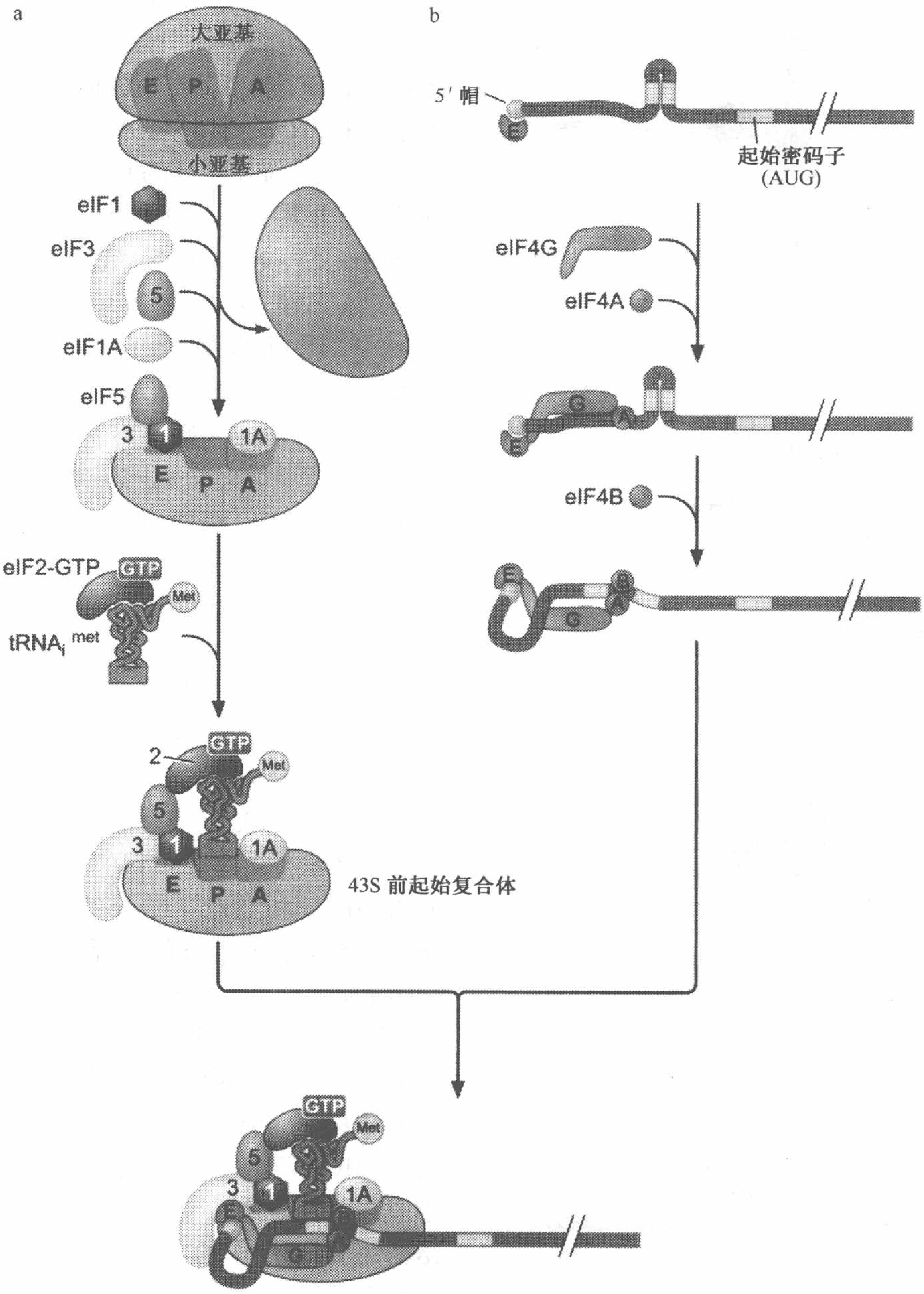


图 14-26　真核细胞核糖体小亚基和起始 tRNA 装配到 mRNA 上。

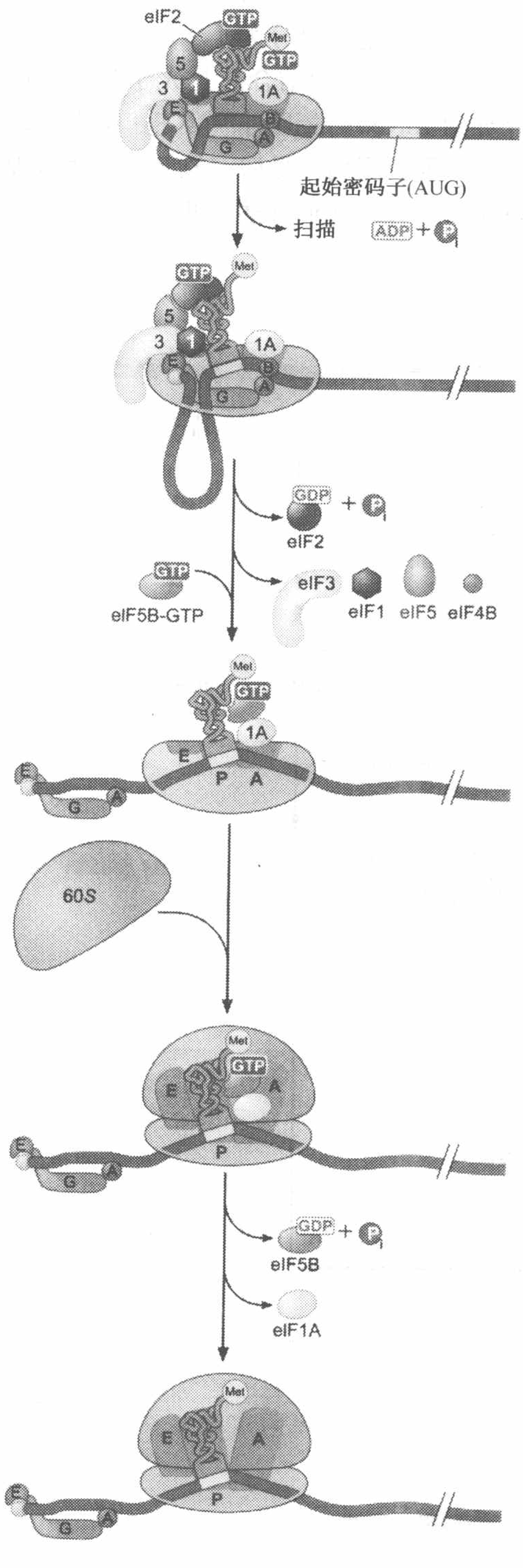


图 14-27 真核细胞核糖体小亚基对起始 AUG 的识别。

真核生物核糖体完成了一个周期的翻译之后，就解体为游离的大小亚基，此时四个起始因子——eIF1、eIF1A、eIF3 和 eIF5——结合于小亚基上。在酵母细胞中，其中的三个因子（eIF1、eIF3 和 eIF5）在与小亚基结合之前先组装成一个多因子复合体；类似的相互作用也很可能发生在其他物种中。这四个因子共同作用，以一种与原核起始因子 IF3 和 IF1 作用相同的方式来防止大亚基和 tRNA 同时与 A 位点结合。起始 tRNA 被三个亚基组成的 GTP 结合蛋白 eIF2 种送至小亚基上。与 IF2 类似，eIF2 以 GTP 结合的形式与起始 tRNA 结合。产生的复合体称为三重复合物（Ternary Complex，TC）。对于真核生物来讲，起始 tRNA 带有甲硫氨酸，而不是 N-乙酸甲硫氨酸，称为 Met-$tRNA_i^{Met}$。eIF2 将 Met-$tRNA_i^{Met}$ 放置在小亚基中将形成 P 位点的位置上，引成 43S 前起始复合物（43S preinitiation complex）。

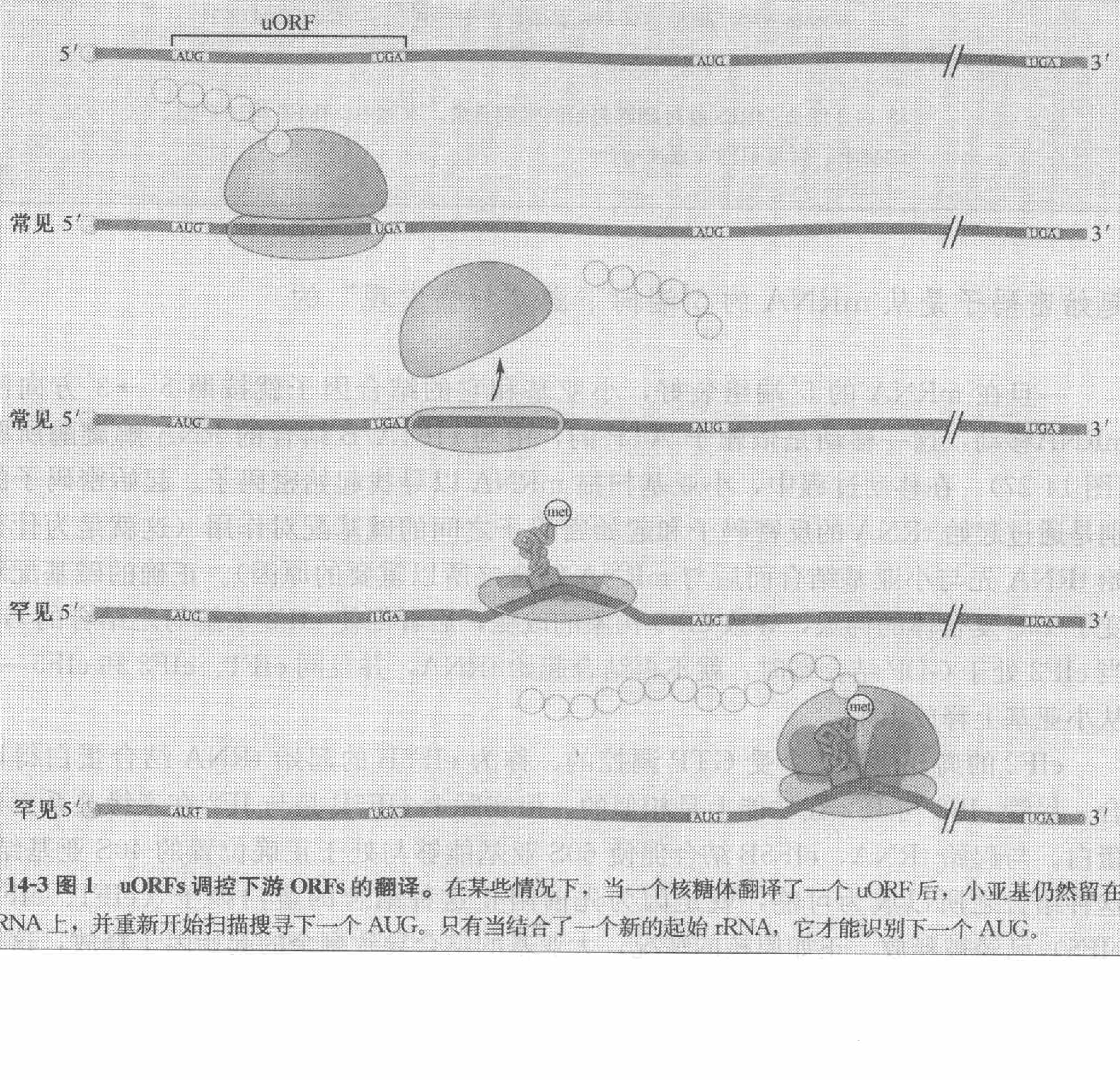
在另外一系列反应中，mRNA 为小亚基对其的识别做好准备。这一过程以帽结合蛋白 eIF4G 识别 5′端帽为起始。而后，一系列附加的起始因子募集进来。eIF4G 同时结合 eIF4E 和 mRNA，而 eIF4A 结合 eIF4G 和 mRNA（图 14-26b）。eIF4G 和 eIF4A 的这种联合尤为重要——细胞中翻译的整体水平由这个步骤中一个叫做 eIF4E 的与 eIF4G 相互竞争结合的蛋白家族来控制（见本章后面部分的翻译调控）。这一复合体而后又加进 eIF4B，后者激活 eIF4A 的 RNA 解螺旋酶活性。解旋酶能够解开任何可能形成于 mRNA 末端的二级结构（如发卡结构）。解除二级结构是非常必要的，因为 mRNA5′端必须消除二级结构才能与小亚基结合。最后，结合在消除了较高级结构的 mRNA 上的 eIF4A 因子与结合在小亚基上的起始因子之间的相互作用将 43S 前起始复合物募集到 mRNA 上，形成 48S

前起始复合物（48S preinitiation complex）。

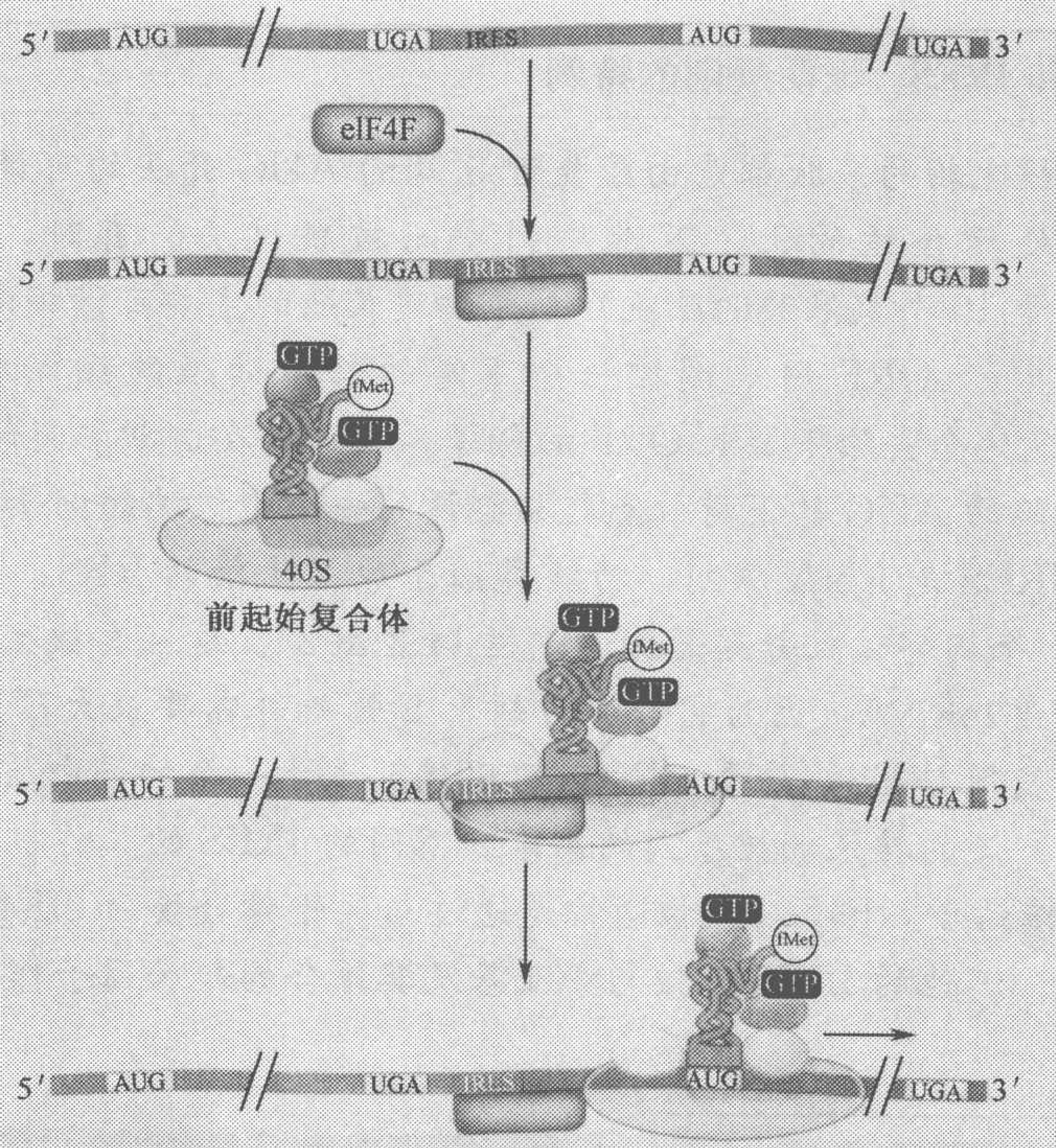
框 14-3　uORF 和 IRES：证明规则的特例

不是所有真核生物的多肽都是由最靠近 5′端的 AUG 开头的可读框编码的。有时，第一个 AUG 不处于合适的上、下游序列中而被跳过去。有时，短的上游可读框（uORF，编码少于 10 个氨基酸的肽段）见于编码长肽的主要可读框的上游（框 14-3 图 1）。在这些情况中，uORF 用来调控一个下游的长可读框的翻译水平。通常，这些 uORF 减少而不是完全去除下游长 ORF 的翻译。我们将在这章举例讨论这一情况。

在最靠近 5′端的 AUG 的下游位点起始翻译的一个更极端的例子是内部核糖体进入位点（IRES）。IRES 是 RNA 序列，功能类似于原核生物的 RBS。它们募集小亚基结合与 mRNA 的一个内部位点并起始翻译（框 14-3 图 2）。这在真核生物转录物中相当少见，多编码于病毒 mRNA，因为它们通常缺乏 5′帽端并需要最大限度地利用它们的基因组序列。通过使用 IRES，病毒 mRNA 可以编码多于一个的蛋白质，减少对每个蛋白质编码序列要有广泛的转录调控序列的需求。不同的 IRES 的工作机制不同。至少有一个病毒 IRES 直接结合 eIF4G，模拟通常是通过与 5′帽结合蛋白 eIF4E 的相互作用来募集该蛋白的方式。其他的 IRES 直接与小亚基及其结合的起始因子作用（如，eIF3）。



框 14-3 图 1　uORFs 调控下游 ORFs 的翻译。在某些情况下，当一个核糖体翻译了一个 uORF 后，小亚基仍然留在 mRNA 上，并重新开始扫描搜寻下一个 AUG。只有当结合了一个新的起始 rRNA，它才能识别下一个 AUG。



框 14-3 图 2　IRES 绕过翻译起始的常规要求。本例中，IRES 绕过 5′帽的要求，而与 eIF4G 直接结合

起始密码子是从 mRNA 的 5′端向下游“扫描发现”的

一旦在 mRNA 的 5′端组装好，小亚基和它的结合因子就按照 5′→3′方向沿着 mRNA移动，这一移动是依赖于 ATP 的，由与 eIF4A/B 结合的 RNA 解旋酶所驱动（图 14-27）。在移动过程中，小亚基扫描 mRNA 以寻找起始密码子。起始密码子的识别是通过起始 tRNA 的反密码子和起始密码子之间的碱基配对作用（这就是为什么起始 tRNA 先与小亚基结合而后与 mRNA 结合之所以重要的原因）。正确的碱基配对改变了 43S 复合体的构象，导致 eIF5 构象的改变，后者促使 eIF2 水解与之结合的 GTP。当 eIF2 处于 GDP 结合态时，就不再结合起始 tRNA，并且同 eIF1、eIF3 和 eIF5 一道，从小亚基上释放出来。

eIF2 的离开使第二个受 GTP 调控的、称为 eIF5B 的起始 tRNA 结合蛋白得以结合。尽管 eIF2 与 IF2 在功能上是相似的，但实际上 eIF5B 是与 IF2 在亲缘关系更近的蛋白。与起始 tRNA，eIF5B 结合促使 60S 亚基能够与处于正确位置的 40S 亚基结合。这种结合之所以成为可能，就是因为先前阻止这种结合的蛋白因子（eIF1、eIF3 和 eIF5）已经被释放。正如原核的情况，大亚基的结合导致剩余的起始因子释放，这是通

过促进 IF2 的类似物——eIF5B 对 GTP 的水解而完成的。

这些事件的结果是，Met-tRNA$_i^{Met}$ 被置于新的 80S 起始复合体（80S initiation complex）的 P 位点。随着起始密码子和 Met-tRNA$_i^{Met}$ 被置于 P 位点，真核细胞的核糖体已经为在 A 位点接受一个负载 tRNA 并开始第一个肽键的合成做好了准备。

翻译起始因子使真核 mRNA 保持环状

除了与 mRNA 的 5′端结合，起始因子通过 poly-A 尾与 mRNA 的 3′端紧密结合（图 14-28）。这是通过 eIF4F 和包绕在 poly-A 尾的 poly-A 尾结合蛋白（poly-A binding protein）之间的相互作用实现的。由于 eIF4F 和 poly-A 尾结合蛋白都在多轮翻译循环中与 mRNA 结合，因此（mRNA 的）5′和 3′端之间存在稳定的相互作用。这种作用在 5′和 3′端之间形成一蛋白质桥，使 mRNA 维持一种环状的构象。poly-A尾能够提高 mRNA 的翻译这一现象早已为人所知。翻译起始因子在 poly-A 尾存在的条件下使 mRNA“环化”的发现为这一现象提供了一个简单、合理的解释：一旦核糖体完成了通过 poly-A 尾而环化的 mRNA 的翻译，新释放的核糖体被理想地置于同一 mRNA 的起始翻译位点上。

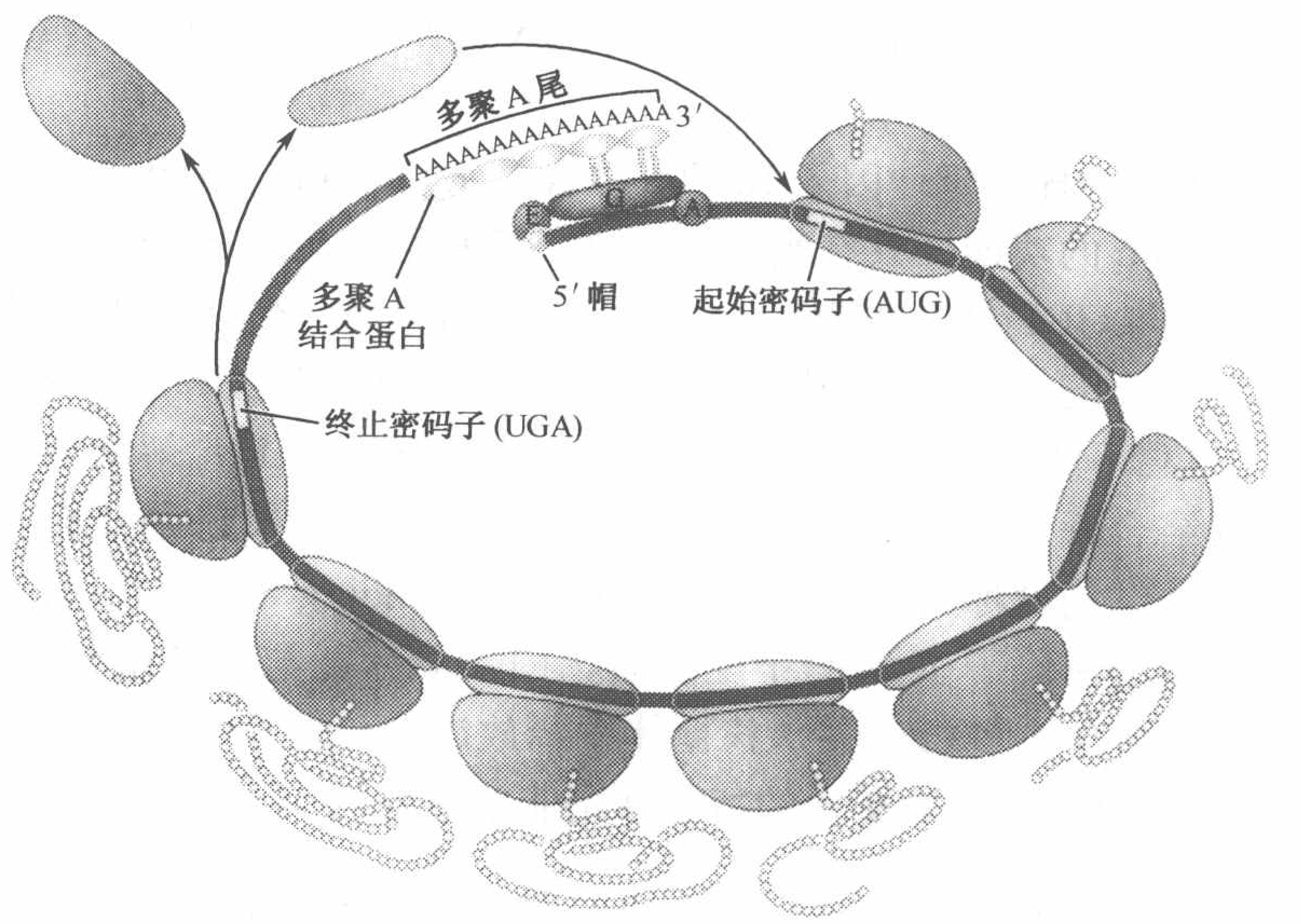


图 14-28 真核 mRNA 环化模型。环化被认为是由 eIF4G 和多聚 A 尾结合蛋白的相互作用介导的。

翻译延伸

一旦核糖体和负载的起始 tRNA 组装于 P 位点上，多肽的合成就能够开始了。为了使氨基酸能正确加入，有 3 个关键的事件是必须发生的（图 14-29）。首先，在 A 位点上的密码子的指导下，正确的氨基酰tRNA置于 A 位点上；第二，A 位点的氨基

酰tRNA与P位点的肽酰 tRNA 上的肽链形成肽键。如我们之前所见，这一肽转移酶（peptidyl transferase）反应导致多肽链从 P 位点的 tRNA 上转移到 A 位点的负载 tRNA 的氨基酸残基上；第三，在 A 位点形成的肽酰 tRNA 和相应的密码子必须易位（translocate）至 P 位点，以使核糖体为下一循环的密码子识别和肽键形成做好准备。同 mRNA 的起始位置必须精确一样，（核糖体在 mRNA 上的）移动也必须准确以保持正确解读的遗传信息可读框。两种已知的延伸因子（elongation factor）调控这些事件。这两种因子都利用 GTP 结合和水解产生的能量来提高核糖体作用的效率和准确性。

与翻译的起始不同，延伸的机制在原核细胞和真核细胞之间是高度保守的。我们将延伸的讨论仅限于原核细胞上，它已经被详细地了解了，但真核细胞的翻译延伸过程无论在涉及的因子和机制上都与原核细胞相似。

氨基酰 tRNA 由延伸因子 EF-Tu 送达 A 位点

氨基酰 tRNA 不能独自与核糖体结合。相反，它们是由延伸因子 EF-Tu“护送”到核糖体上的（图 14-30）。一旦 tRNA 被氨基酰化，EF-Tu 就结合于 tRNA 的 3′端，覆盖耦合的氨基酸。这个作用防止了氨基酰 tRNA 参与肽键的形成，直到从 EF-Tu 中释放出来。

与起始因子 IF2 一样，延伸因子 EF-Tu 结合并水解 GTP，并且鸟嘌呤核苷酸结合的类型决定了 EF-Tu 的功能。只有当 EF-Tu 与 GTP 结合后，EF-Tu 才能结合氨基酰 tRNA。而 EF-Tu 与 GDP 结合或EF-Tu单独时，与氨基酰 tRNA 没有什么亲和力。因此当 EF-Tu 水解结合的 GTP 时，其所结合的氨基酰 tRNA 就被释放出来。结合氨基酰 tRNA 的EF-Tu单独不能有效地水解 GTP。相反，EF-Tu GTP 水解酶活性是在与大亚基的一个结构域结合而得以激活的，该结构域在大亚基加入起始复合体时激活 IF2 GTP 水解酶。这一结构域称为因子结合中心（factor binding center）。只有当 tRNA 置于 A 位点并且正确的密码子-反密码子配对后，EF-Tu 才能与因子结合中心相互作用。然后，EF-Tu 水解结合的 GTP，并从核糖体中释放出来（图 14-30）。正如下面将讨论的，EF-Tu 调控 GTP 的水解对翻译的特异性是非常重要的。

核糖体利用多种选择机制排斥错误配对的氨基酰 tRNA

翻译的错误率为 10^{-3}～10^{-4}。这就是说，在合成的蛋白质中，每 1000 个氨基酸中只有不多于一个氨基酸是错误的。选择正确的氨基酰 tRNA 的最终基础在于氨基酰 tRNA 和核糖体 A 位点的密码子之间的碱基配对。尽管如此，正确的密码子-反密码子配对和不正确的配对之间的能量差异尚不能完全解释这样高的准确率。在许多情况下，密码子-反密码子相互作用的三对核苷酸中只有一对是不配对的，但核糖体很少允许这样的氨基酰 tRNA 继续翻译过程。至少有 3 种不同的机制促成了翻译的高准确率（图 14-31）。这些机制都选择排斥密码子-反密码子配对错误（的氨基酰 tRNA）。

其中一个有助于密码子识别忠实性的机制涉及小亚基 A 位点内 16*S* RNA 的两个

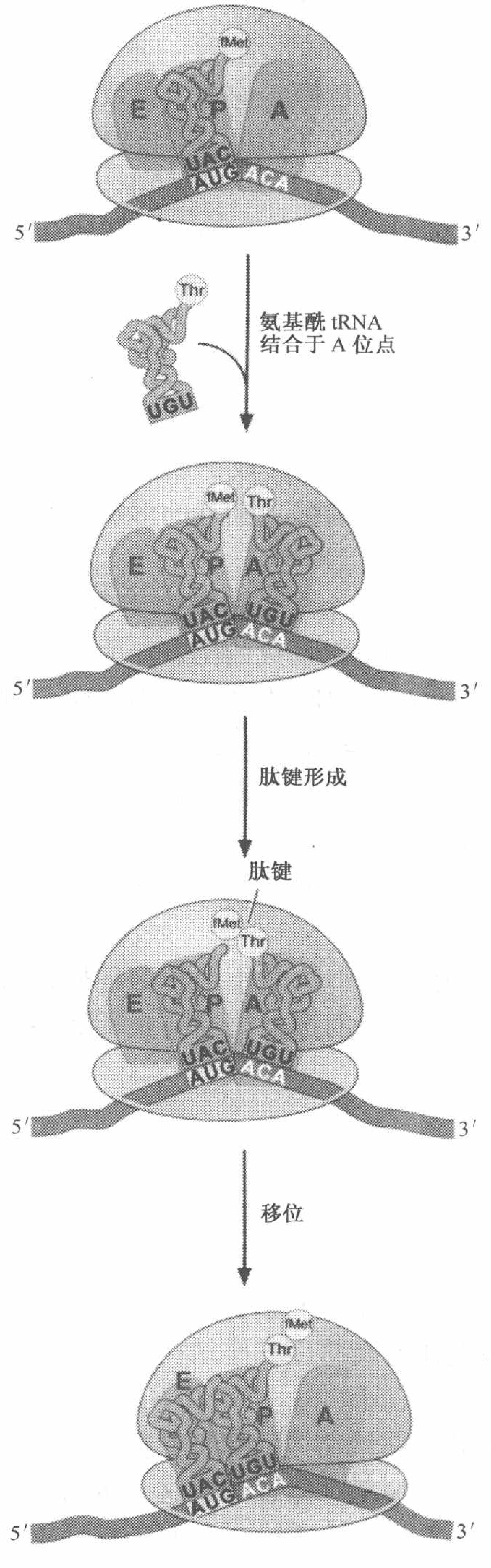


图 14-29　翻译的几个步骤。

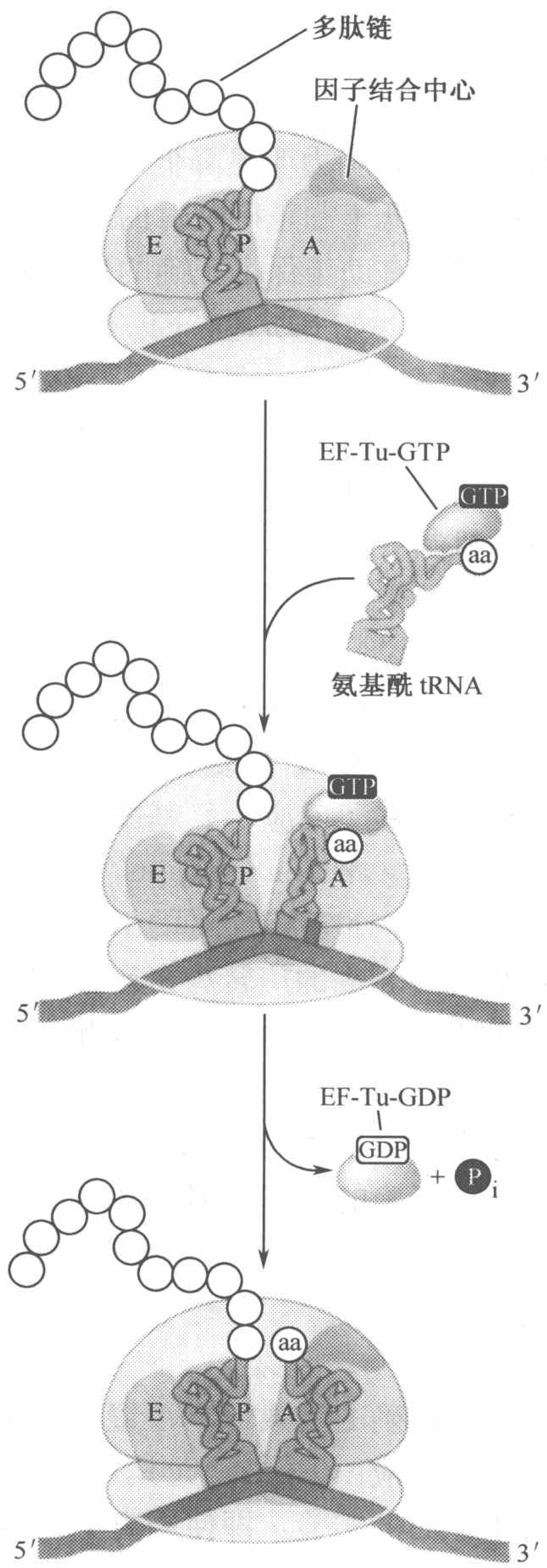


图 14-30　EF-Tu 护送氨基酰 tRNA 至核糖体的 A 位点。 当负载 tRNA 最初与核糖体的 A 位点发生作用时，它们是与 EF-Tu-GTP 结合在一起的。当正确的密码子-反密码子相互作用发生时，EF-Tu 与因子结合中心发生反应，水解所结合的 GTP，并从 tRNA 和核糖体上脱离。

相连的腺嘌呤残基。这两个腺嘌呤与反密码子和密码子的前两个碱基之间每个正确配对所形成的小沟形成氨基酸（图 14-31a）。正如你会想到的（图 6-10），在 Watson-Crick

的G：C和A：U小沟中的氢键特性是配对非常相似的。因此，相邻的16S RNA的A残基不能辨别G：C或A：U配对，因而认为任何一对都是正确的。相反，非Watson-Crick碱基配对所形成的小沟就不能被（16S RNA的）这两个腺嘌呤识别，导致对非正确tRNA的亲和力明显降低。这些作用的净结果是，正确配对的tRNA从核糖体解离的速度要远低于非正确配对的tRNA。

第二个有助于保证正确的反密码子-密码子配对的机制涉及EF-Tu的GTP酶活性（图14-31b）。如上所述，EF-Tu从tRNA的释放需要GTP的水解，这一作用对正确的反密码子-密码子配对是非常敏感的。即使只有一个碱基的不配对也将改变EF-Tu的位置，降低其与因子结合中心的相互作用能力。这就导致了EF-Tu的GTP酶活性的急剧下降。这一机制是动力学敏感性的一个范例，并且与DNA合成过程中保证正确的碱基配对的机制有关（第8章）。在这两个过程中，正确的碱基配对的形成都明显地加快了一个重要的生化步骤的速度。对DNA聚合酶来说，这一步是磷酸二酯键的形成。对蛋白质合成来说，这一步就是EF-Tu对GTP的水解。

第三个保证碱基配对正确性的机制是EF-Tu释放后的一个校正机制。当负载tRNA与EF-Tu-GTP的复合体进入A位点时，它的3′端远离肽键形成的位点。为了成功地进行肽转移酶反应，氨基酰tRNA必须旋转进入到大亚基的肽转移酶中心，这一过程称为入位（accommodation，图14-31c）。在八位过程中，氨基酰tRNA的3′端移动了几乎70Å。非正确配对的tRNA在入位过程中经常从核糖体上脱离下来。有假设认为氨基酰tRNA的旋转为密码子和反密码子的作用带来了张力，而只有正确配对的反密码子才能承受这种张力。因此，错配的氨基酰tRNA就更有可能在肽转移酶反应进行之前从核糖体上脱离下来。

总之，除了密码子-反密码子的相互作用，核糖体还利用小沟的作用和两种校正机制来保证正确的氨基酰tRNA结合在A位点上。这三种额外的选择机制阻止了那些不能形成正确的密码子-反密码子作用的氨基酰tRNA的留存。

核糖体是核酶

一旦正确的负载tRNA置于A位点并旋转进入肽转移酶中心，肽键的形成就开始了。这一反应由RNA，尤其是大亚基的23S rRNA组分来催化。这种论点的早期证据来自于下面的实验：即使剥离了很多蛋白质的大亚基依然能够进行肽键的形成。对肽转移酶其实由RNA组成的证明来自于一个原核生物核糖体的高精度三维结构，它揭示没有一个氨基酸位于活性位点的18Å之内（图14-32）。

近来对*E. coli*核糖体与mRNA及tRNA整合复合物三维结构的解析已揭示出，一个蛋白（L27）的氨基末端确定接触到了活性部位。这一发现可能意味着这一蛋白在催化过程中起到某种作用。为了检测这种可能性，将L27蛋白氨基端9个处于活性中心附近的氨基酸进行缺失突变。生成的细胞所产生的核糖体带有降低的、但可检测到的肽酰转移酶活性，这就清楚地说明L27蛋白的这一区域参与肽酰转移酶的活性构成。然而突变的核糖体依然以野生型核糖体蛋白合成水平的30%～50%来合成蛋白，并且含有突变核糖休的细胞仍然可以生长分裂。与底物（氨酰-tRNA）在溶液中单独存在时相

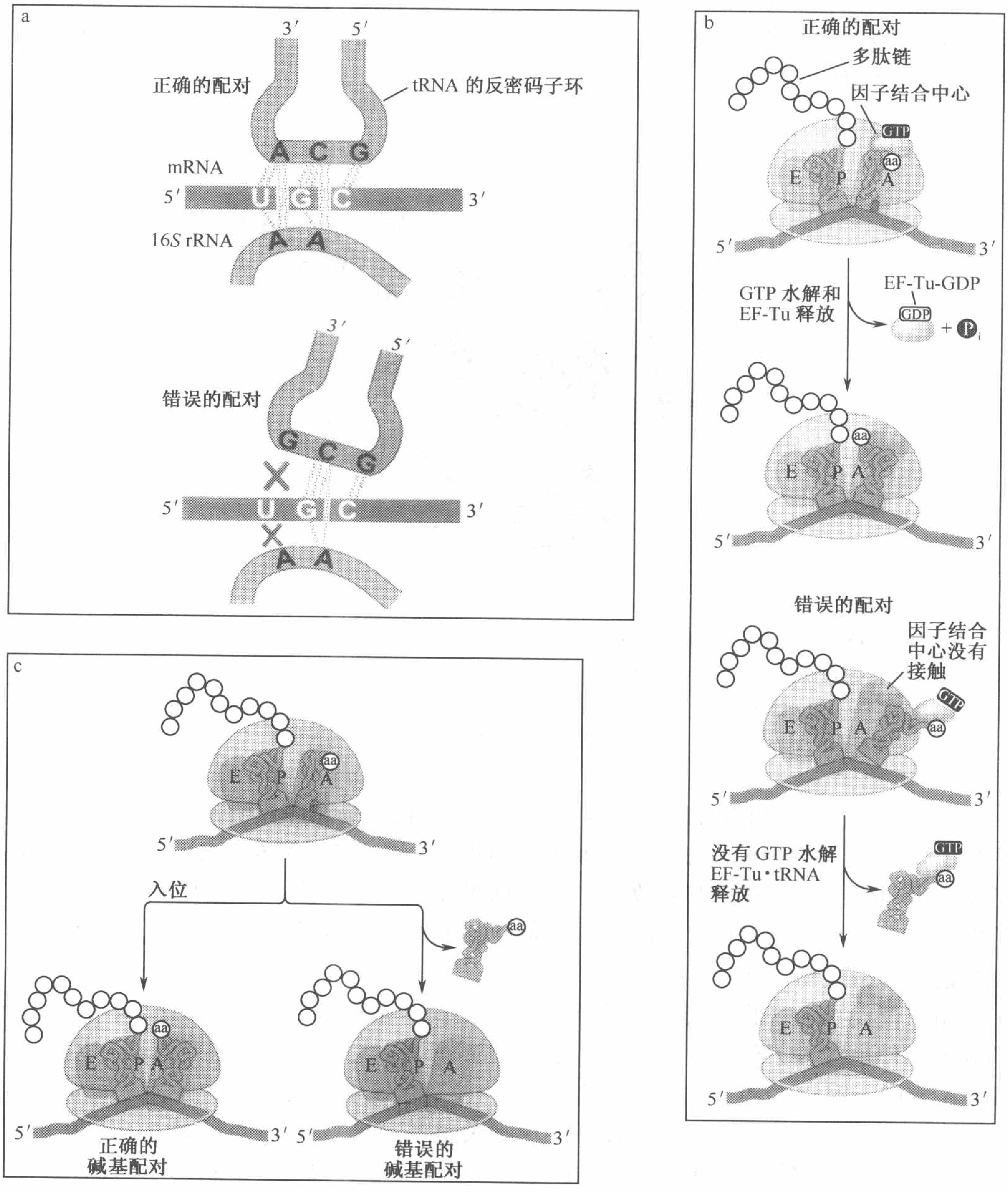


图 14-31 保证 rRNA 和 mRNA 正确配对的三种机制。 a. 只有当密码子-反密码子正确配对时，16S rRNA 的两个腺嘌呤残基与配对碱基对的小沟之间才能形成额外的氢键。b. 正确的碱基配对使得结合氨基酰 tRNA 的 EF-Tu 与因子结合中心相互作用，诱发 GTP 的水解和 ET-Tu 的释放。c. 只有碱基配对正确的氨基酰 tRNA 在肽键形成过程中旋转进入正确位置时，才能保持与核糖体的结合。这一旋转易位称为 tRNA 入位。

比，核糖体提高了 10^7 倍的肽键形成速率。显然，即使 L27 不在活性位点上，依然保留了绝大多数的速率提升。因此，尽管这一蛋白促进肽键形成，但它对于肽酰转移酶活性并不是必需的。像其他核糖体蛋白一样，L27 最有可能的功能就是将一个或更多的

RNA 组分正确地摆放到活性位点。更为重要的是，因为这个蛋白是唯一一个距离（活性中心）最近来行使催化功能的蛋白，所以核糖体的 rRNA 组分一定承担着催化肽键形成的主要功能。

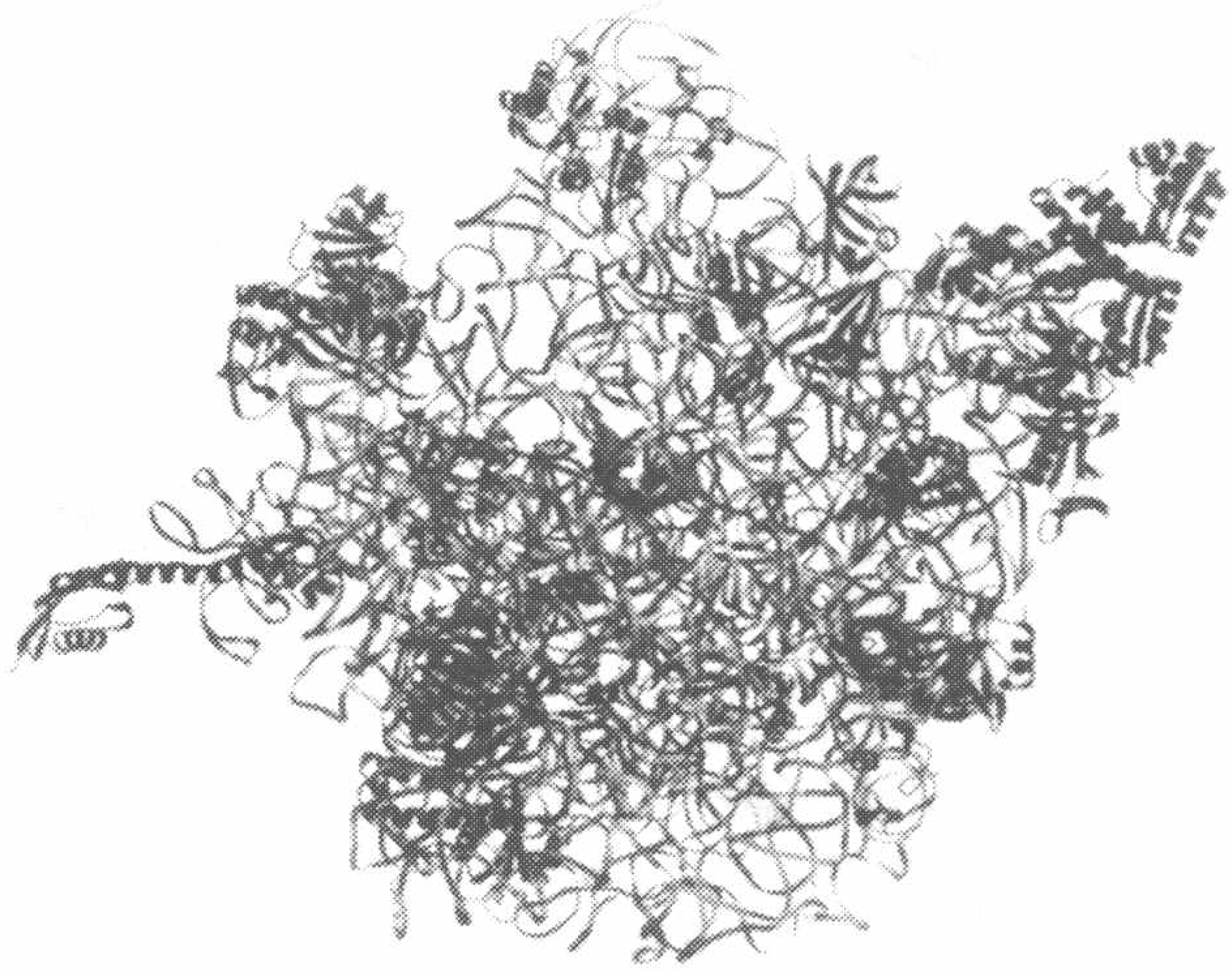


图 14-32　RNA 环绕着大亚基的肽转移酶中心。所示为细菌 50S 亚基的三维结构。rRNA 为灰色，核糖体蛋白质为紫红色。紧邻肽酰转移酶中心的 A 和 P 位点的 tRNA 的 3′端分别以绿色和红色表示。(Yusupov M. M. et al. 2001. Science 292：883-896.）图像使用 MolScript、BobScript 和 Raster 3D 软件制作。

23*S* rRNA 是如何催化肽键形成的呢？准确的机制仍然需要研究，但某些答应已经清楚了，首先 23*S* rRNA 与处于 A 和 P 位点的 tRNA 的 CCA 末端之间的碱基配对，帮助氨基酰 tRNA 的 α 氨基基团攻击结合于肽酰-tRNA 的多肽的碳基团。这一作用也有可能稳定入位后的氨基酰 rRNA。这类催化机制称为镝催化（centropic catalysis）。也就是说，酶将底物带到一起来激发催化。

由于仅仅靠底物的紧密相邻并不足以产生高水平的催化作用，因此很可能 rRNA 的其他元件也参与了催化。实际上，去掉 23*S*rRNA 上一个高度保守残基（*E. coli*23*S* rRNA 上的 A2451 位点）的 2′-OH 这样的一种变化使催化速率至少降低了十倍。近期的研究又发现了一个未曾料到的在催化中很关键的 RNA：P 位点 tRNA。去掉 P 位点 tRNA3′端 A 残基的 2′-OH 导致催化速率降低了 10^6 倍。这种“底物辅助的催化”是一个尤为有趣的发现，因为这预示着氨酰-tRNA 本身就携带着关键的催化元件。这一发现意味着，在核糖体进化之前，tRNA 就已经为催化其自身蛋白质的合成提供关键元件了。

基于一系列相关考虑，人们提出，P 位点 tRNA 的 2′-OH 可能充当“质子穿梭机制”的一部分（图 14-33）。在这一模型中，2′-OH 向氨酰-tRNA 的 3′-OH 提供一个 H，并从发起攻击的连接于 A 位点 tRNA 的氨基酸的 α-氨基基团得到一个质子。十分重要的是，这两个发现都强烈支持这样一种假设，即正是 RNA 而不是蛋白质催化了肽键的形成。尽管如此，关于核糖体怎样催化肽键形成仍有很多未知。

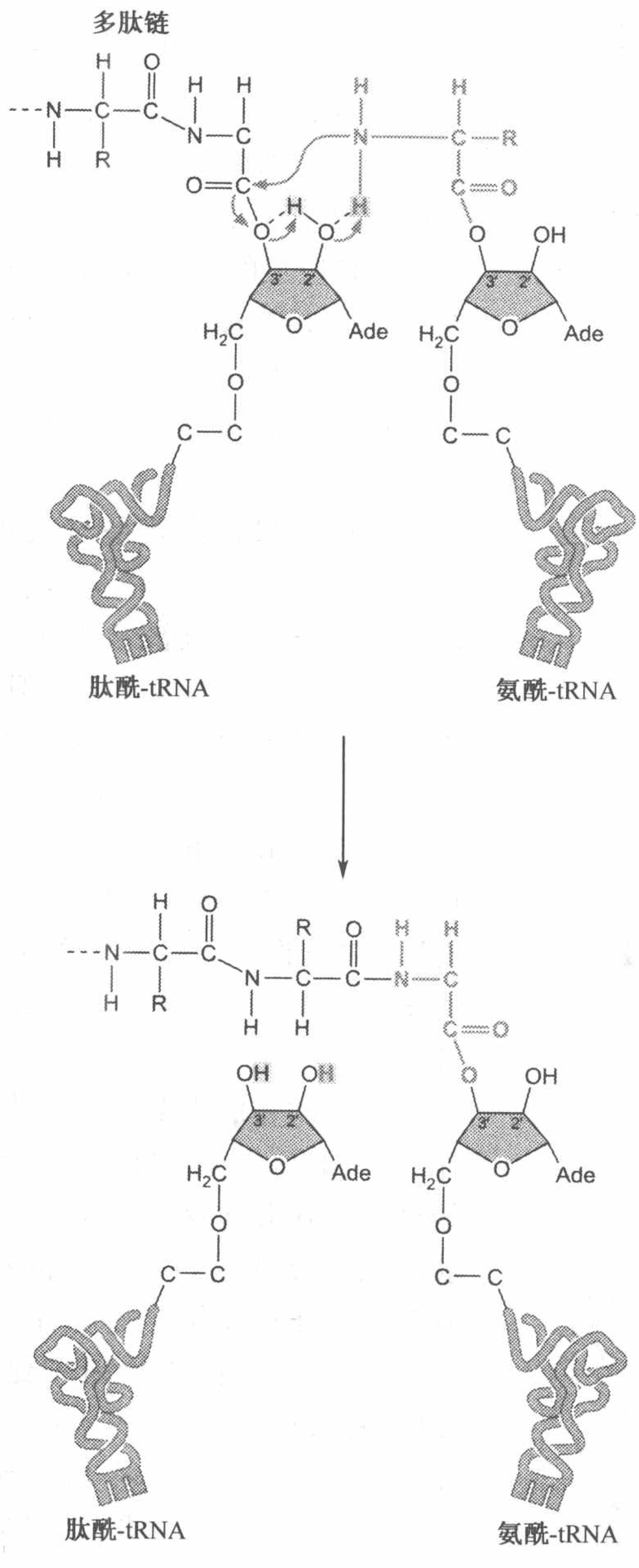


图 14-33　肽键形成中 P 位点 tRNA2′-OH 的功能。肽酰-tRNA 最后一个“A”的 2′-OH 对于肽键形成十分关键。基于以上发现，有人提出 2′-OH 的氢组分可能参与如图所示的一个“质子穿梭机制”。在此模型中，随着肽酰-tRNA 与肽链之间的键断裂，3′O 从 2′-OH 夺取一个 H（以黄色表示），而 2′O 反过来从攻击羧基的氨基基团夺取一个 H（以绿色表示）。红色箭头显示所提出的在肽键形成过程中电子的运行方向。

肽键形成和延伸因子 EF-G 促进了 tRNA 和 mRNA 的易位

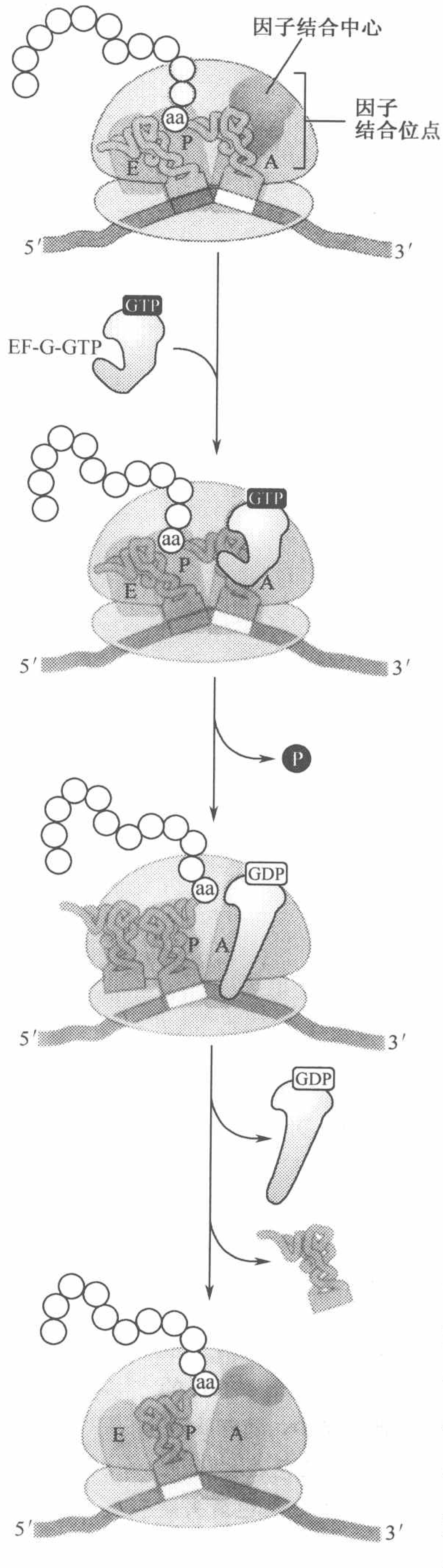


图 14-34 EF-G 刺激易位需要 GTP 的水解。

一旦肽转移酶反应开始发生，P 位点的 tRNA 就被去氨基（不再结合氨基酸），而多肽链则连接到 A 位点的 tRNA 上。要使肽链延伸的新一轮循环得以发生，P 位点的 tRNA 必须移至 E 位点，而 A 位点的 tRNA 必须移至 P 位点。同时，mRNA 也必须移动 3 个核苷酸，以使下一个密码子暴露出来。这些移动是在核糖体内协调进行的，统称为易位（translocation）。

易位的起始步骤是与肽转移酶反应耦联的（图 14-34）。一旦多肽链移至 A 位点的 tRNA，A 位点和 P 位点的 tRNA 就被认为进入了杂合状态。A 位点 tRNA 的 3′端与多肽结合，但其反密码子仍然与解码中心 A 位点的密码子配对。相似地，这时已脱氨基的 P 位点 tRNA 不再与多肽结合，但是它的反密码子仍然与 P 位点的密码子配对。因此，大亚基的易位是先于小亚基的易位，而 tRNA 就处于“杂合状态”。它们的 3′端移入一个新的位置，但它们的反密码子末端仍位于未发生肽转移酶反应时的位置上（图 14-34，第一步）。

易位的完成需要第二个称为 EF-G 的延伸因子的作用。EF-G 只有在和 GTP 结合的情况下才能结合核糖体。在肽转移酶反应发生后，P 位点 tRNA 缺少肽的结合及 A 位点 tRNA 的移动就在 A 位点的大亚基部分为 EF-G 结合腾出了位点。当 EF-G-GTP 结合后，它与大亚基的因子结合中心相作用，刺激 GTP 的水解。GTP 的水解改变了 EF-G-GDP 的构象，使它可以进入小亚基并刺激 A 位点 tRNA 的易位（图 14-34，第三步）。当易位完成后，核糖体构象的改变极大地降低了其对 EF-G-GDP 的亲和力，使得这一延伸因子从核糖体中释放出来。这些事件一起导致 A 位点的 tRNA 易位至 P 位点上，P 位点的 tRNA 易位至 E 位点，以及 mRNA 移动正好 3 个核苷酸（图 14-34，第四步）。此时，核糖体为新一轮的氨基酸加入的开始做好了准备。

EF-G 通过取代结合在 A 位点的 tRNA 来驱动易位

EF-G 促使易位的精确机制目前尚不完全清楚，但其中的部分机制与 EF-G-GDP 占据 A 位点的解码中心的能力有关。通过与解码中心的相互作用，EF-G-GDP 取代 A 位点的 tRNA 使后者转移至 P 位点。就像多米诺现象，A 位点的 tRNA 进入 P 位点意味着 P 位点的 tRNA 必须移到 E 位点。在 tRNA 的移动过程中，mRNA 也移动了 3 个核苷酸的位置。mRNA 的移动是由移动的 A 位点的 tRNA 和 mRNA 之间的碱基配对介导的，碱基配对在易位过程中一直维持着。基本上，mRNA是被移动的 A 位点 tRNA 拉着走。的确，罕见的 4 个核苷酸的反密码子的“移码”tRNA（因此可以弥补某些移码突变）移动 mRNA 4 个核苷酸，而不是 3 个。与 A 位点 tRNA 的移动相反，P 位点的 tRNA 移至 E 位点打断了 tRNA 和 mRNA 之间的这种碱基配对。这样，现在 E 位点上卸载的（脱氨基酸的）tRNA 可以自由地从核糖体中脱离出来，并通过氨基酰 tRNA 合成酶重新装载氨基酸。

核糖体亚基的变化也对易位有所贡献。例如，在易位时，小亚基相对于大亚基发生了逆时针转动，另外，小亚基结构必须发生变化以使 EF-G-GDP 在易位结束后释放出来。结构研究还发现在 A、P 和 E 位点之间有“门”将它们分开。因此，为了使 tRNA 移至新的位置，这些“门”必须打开。EF-G-GDP 占据 A 位点的解码中心保证了在易位之后，先前的 A 位点 tRNA 不能再移回到 A 位点。当 EF-G 释放后“门”可能就又关闭了。

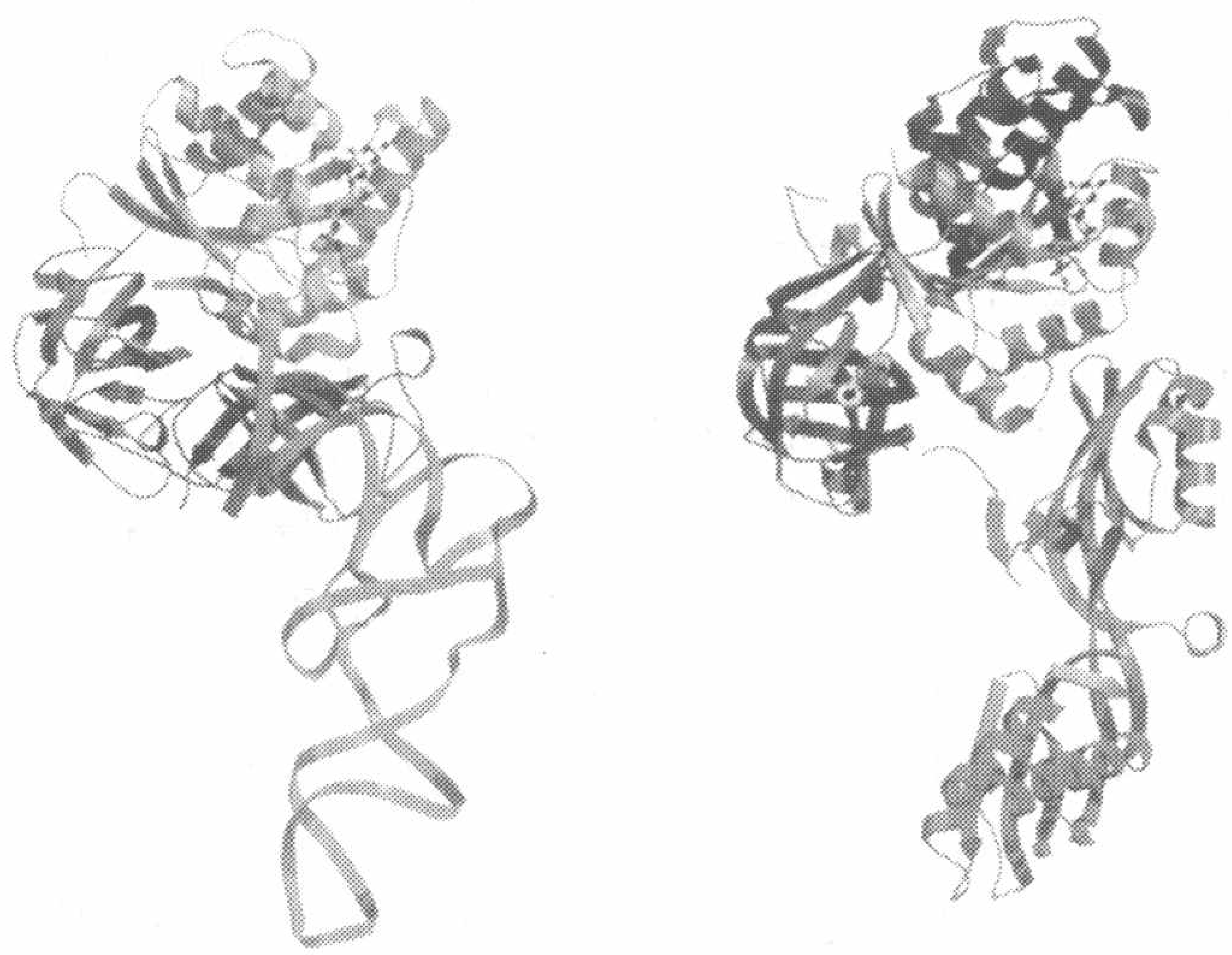


图 14-35 延伸因子的结构比较。左图为 EF-Tu-GDPNP-Phe-tRNA，右图为 EF-G-GDP。GDPNP 为不能水解的 GTP 类似物，在决定三维结构时用于在 GTP-结合的构象中锁定分子。请注意 EF-G 中示以绿色标注的结构和与 EF-Tu 结合的 tRNA（也以绿色表示）的结构相似性。（左图：Nissen P. et al. 1995. Science 270：1464-1472. 右图：al-Karadaghi S. et al. 1996. Structure 4：555-565.）图像使用 MolScript、BobScript 和 Raster 3D 软件制作。

EF-G-GDP 与 A 位点的解码中心的作用何以如此有效？EF-G 和 EF-Tu-tRNA 的晶体结构为这一问题提供了明确的答案。EF-G-GDP 和 EF-Tu-GTP-tRNA 具有非常相似的结构（图 14-35）。回想 EF-Tu-GTP-tRNA 也结合于 A 位点的解码中心。这一结构相似性的最显著特征在于，仅管 EF-G 仅由单一的多肽链组成，它的结构却模仿了与蛋白质结合的 tRNA。这是“分子模仿”的一个例子，其中蛋白质以类似 tRNA 的形状出现以促进其与同样的结合位点相结合。有意思的是，关于真核细胞中 EF-G 的同类物（称为 eEF-2）的结构研究发现这一蛋白存在两个完全不同的构象（其中之一结合于抗菌素类壳菌素）。一个构象与 EF-G 的结构（图 14-35）相似，而另一个构象在 tE-NA 模拟区向 GTP 结合区显著位移后产生。可能这样一种构象上的变化对 EF-G 在位移过程中行使功能是十分重要的。

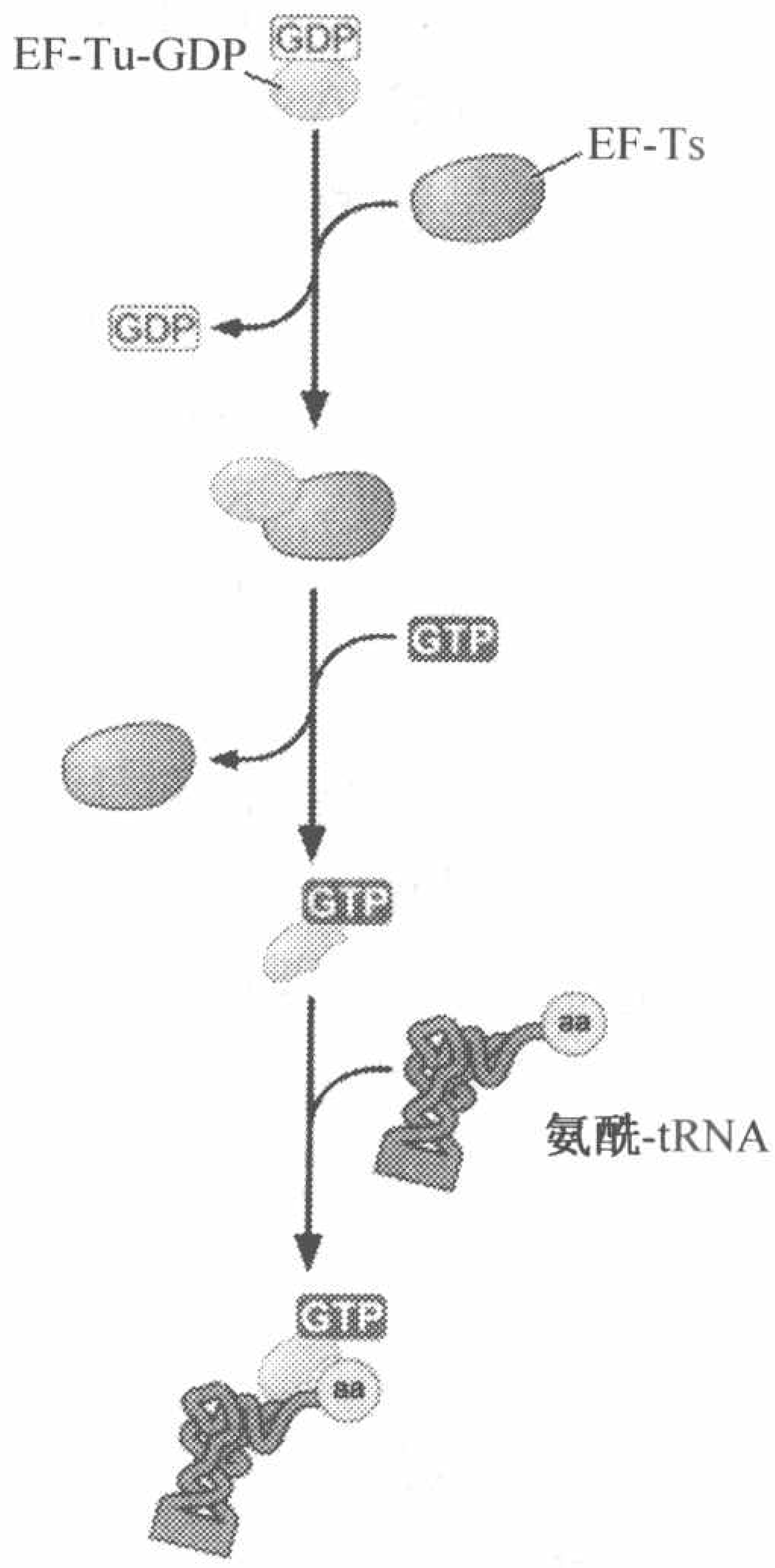


图 14-36 EF-Ts 促进 GDP 与 EF-Tu 的解离。结合于 EF-Tu 的 GDP 释放非常缓慢。EF-Ts 结合于 EF-Tu-GDP，引发 GDP 的快速解离。结合于 EF-Tu-EF-Ts 复合物中的 EF-Tu 上的 GTP 取代 ET-Ts 并释放 EF-Tu-GTP。EF-Tu-GTP 随后能够结合一个新的氨酰-tRNA 并运往核糖体。

EF-Tu-GDP 和 EF-G-GDP 在参加新一轮延伸之前必须将 GDP 换成 GTP

EF-Tu 和 EF-G 是催化蛋白质，在每一轮 tRNA 结合核糖体、肽键形成和易位的过程中只能使用一次。GTP 水解后，这两个蛋白质必须释放所结合的 GDP，并结合一个新的 GTP 分子。对 EF-G 来说，这是一个简单的过程，因为 GDP 与 EF-G 的亲和力要低于 GTP。因此 GTP 水解后 GDP 和磷酸释放出来，游离的 EF-G 很快与一个新的 GTP 分子相结合。而对 EF-Tu 而言，需要另一个蛋白质来完成 GDP 到 GTP 的交换。延伸因子 EF-Ts 是 EF-Tu 的 GTP 交换因子（GTP exchange factor）（图 14-36）。EF-Tu-GDP 从核糖体释放出来后，EF-Ts 与 EF-Tu 结合，促使 GDP 脱落。随后，GTP 结合于 EF-Tu-EF-Ts 复合体，导致它分解为游离的 EF-Ts 和 EF-Tu-GTP。最后，EF-Tu-GTP 结合一个负载的 tRNA，重新形成 EF-Tu-GTP 氨基酰-tRNA 复合体，再一次准备将新的负载 tRNA 送至核糖体。

形成肽键的一个循环要消耗两个 GTP 分子和一个 ATP 分子

我们以计算能量消耗来结束对延伸过程的讨论。有多少三磷酸核苷酸分子在每一轮的肽键形成中被消耗了呢（不算氨基酸生物合成的能量消耗和起始与结束阶段的消耗）？还记得一个 ATP 分子被氨基酰 tRNA 合成酶消耗，形成了连接氨基酸和 tRNA 的高能酰基键。这一高能酰基键断裂（所释放的能量）驱动了产生肽键的肽转移酶反应。第二

个三磷酸核苷酸（GTP）分子在 EF-Tu 将负载 tRNA 运送至核糖体的 A 位点和确保正确的密码子-反密码子的识别过程中被消耗，最后，第三个三磷酸核苷酸（GTP）在 EF-G 介导的易位过程中被消耗。这样，一个肽键的形成就消耗了细胞两个 GTP 分子和一个 ATP 分子，翻译延伸过程的每一步消耗一个三磷酸核苷酸。有趣的是，这三个分子中，只有一个（ATP）是与肽键的形成有关。另外两个（GTP）分子的能量用于保证翻译的准确和次序（框 14-4）。

在翻译延伸过程的讨论中，我们并没有区分原核细胞和真核细胞。虽然真核细胞中与 EF-Tu（eEF1）和 EF-G（eEF2）相对应的因子的名称不同，但它们与原核细胞的对应因子的作用是非常相似的。

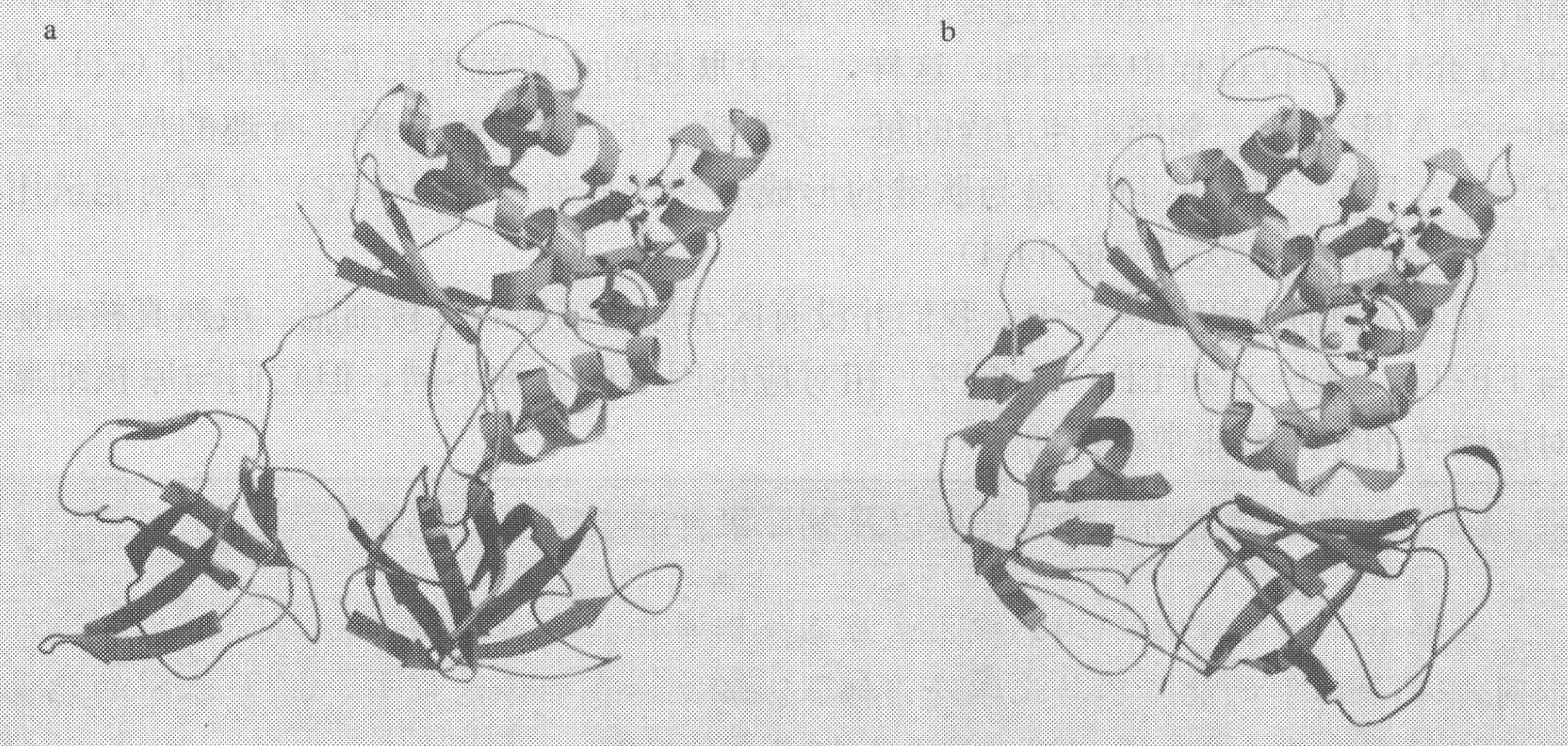
框 14-4　GTP 结合蛋白、构象转变以及翻译事件的忠实性和次序

在整个翻译过程中，GTP 被用来控制关键事件。与 ATP 在连接氨基酸到 tRNA 上不同，GTP 水解的能量不与化学修饰耦联。相反，在翻译过程中，GTP 水解的能量被用来控制事件的次序和忠实性。这是如何完成的呢？

GTP 结合蛋白参与翻译的一个关键特征是它们的构象因结合哪种鸟苷酸（如 GDP 还是 GTP）而改变。这在框 14-4 图 1 的 EF-Tu 中可见，该图显示了结合 GTP 或 GDP 的 EF-Tu 的三维结构。当与 GTP 结合时，EF-Tu 经历一个重大的构象改变，导致其 tRNA 结合位点的形成。具体来说，取决于结合的核苷酸，EF-Tu 的一个结构域（在框 14-4 图 1 中以洋红色显示）相对于该蛋白质的其他结构域可以进行位置的迁移。结构域位置的这个改变和其他两个结构域的构象改变（以绿松色和深蓝色显示）导致 EF-Tu 上形成一个新的表面用来紧密结合负载 tRNA（可在图 14-35 中看到结合了一个 tRNA 的 EF-Tu）。这样，取决于结合的鸟苷酸形式，这些因子可以有不同的功能，或结合到不同的蛋白质/RNA 上。例如，EF-Tu-GTP 可以结合到氨基酰 tRNA 上，但 EF-Tu-GDP 不能。

在翻译中通过将 GTP 水解与关键事件的完成耦联，可以牢牢掌制这些事件的次序。对 EF-Tu 来说，依赖 GTP 的 EF-Tu 与氨基酰 tRNA 的结合保证了肽键的形成不会在正确的密码子-反密码子配对之前发生。正确碱基配对的形成激发了 GTP 水解。一旦与 GDP 结合，EF-Tu 就从氨基酰 tRNA 上释放出来，使肽键的形成能随后发生。

每个 GTP 调节的辅助蛋白活化 GTP 水解的机制是一样的。在每种情况下，GTP 酶的活性是通过与称为因子结合中心（factor binding center）的大亚基的一个特定区域的相互作用而激发的。这个相互作用因缺乏足够的亲和力，在分离状态不发生。相反，每个 GTP 控制的翻译因子必须与核糖体发生几个其他的重要相互作用来稳定与因子结合中心的精确关联，引发 GTP 酶的活性。实际上，如同我们从 EF-Tu 所看到的，这个相互作用对 EF-Tu、氨基酰 tRNA、mRNA 和核糖体之间相互作用的确切性质非常敏感。因此，与因子结合中心的相互作用监控这些蛋白质和 RNA 与核糖体的所有其他的相互作用。只有当正确的密码子-反密码子配对形成，GTP 结合位点才能够有效地与因子结合中心作用，促使 GTP 水解和蛋白质构象相应改变。



框 14-4 图 1 比较结合 GDP 及结合 GTP 的 EF-Tu。 a. 结合 GDP 的 EF-Tu；b. 结合 GTP 的 EF-Tu。GTP 结合域以松绿色显示。当结合 GTP 时，洋红结构域的旋转以及松绿色和蓝色结构域的改变导致强的 tRNA 结合位点的形成（图 14-35）。[来源 a. Polekhina G. et al，1996. Structure 4：1141—1151. b. Kjeldgard M. et al. 1993，Structure 1：35-50.] 图像使用 MolScript、BobScript 和 Raster 3D 软件制作。

翻译过程对 GTP 的利用与滑动夹装载器（第 8 章，框 8-4）对 ATP 的利用类似。回想在那个例子中，ATP 的结合是将滑动夹与一个起始复合物组装起来所必需的，但 ATP 水解和滑动夹的释放只有在夹子装载器与引物：模板结合处结合后才能发生。在翻译过程中，GTP 是核糖体（在某些情况中是其他的 RNA 和蛋白质）与 GTP 调控因子开始结合的所必需的，GTP 水解只在因子正确地与核糖体作用后才会发生。如同滑动夹的例子，GTP 水解一般导致因子从核糖体上释放出来。

翻译终止

释放因子在终止密码子的作用下终止翻译

氨基酰 tRNA 的结合、肽键的形成和易位的核糖体循环在 3 个终止密码子其中之一进入 A 位点时就终止了。最初是假定有一种或几种可识别这些终止密码子的链终止 tRNA。但是，事实并非如此。终止密码子其实是被称为释放因子（release factors，RF）的蛋白质识别的，这些因子激活多肽链从肽酰 tRNA 中释放的水解反应。

有两类释放因子：Ⅰ类释放因子识别终止密码子，并催化多肽链从 P 位点的 tRNA 中水解释放出来。原核细胞有两种Ⅰ类释放因子：RF1 和 RF2。RF1 识别终止密码子 UAG，RF2 识别终止密码子 UGA，两者皆可识别第三个终止密码子 UAA。真核细胞

只有一种能识别 3 个终止密码的Ⅰ类释放因子：eRF1。Ⅱ类释放因子在多肽链释放后刺激Ⅰ类因子从核糖体中解离开来。原核细胞和真核细胞都只有一种Ⅱ类因子，分别是 RF3 和 eRF3。与 EF-G、EF-Tu 和其他翻译因子一样，Ⅱ类释放因子也是由 GTP 的结合与水解调节的。

Ⅰ类释放因子的一小段区域识别终止密码子并催化多肽链的释放

释放因子是如何识别终止密码子的呢？由于释放因子完全是由蛋白质组成的，因此对密码子的识别必然通过蛋白质-RNA 相互作用。将 RF1 和 RF2（有不同的终止密码子特异性）的编码区序列交换的遗传学实验查明了释放因子的识别区域是串联的 3 个氨基酸。RF1 和 RF2 之间这 3 个氨基酸的交换导致对终止密码子识别特异性的转换。因此这 3 个氨基酸序列称为肽反密码子，必定作用于并识别终止密码子。结合 RF1 的核糖体三维结构证实了 RH 结合于核糖体的 A 位点（图 14-37a）。在该结构中肽反密码子区就位于反密码子（原文为反密码子）附近，但很可能还有其他蛋白区域也对密码子识别有所贡献（图 14-37b）。

Ⅰ类释放因子中促使多肽链释放的区域也被确认了。所有的Ⅰ类因子共有一段保守的三氨基酸（GGQ）序列，这一序列对多肽链的释放是关键的。而且，结合 RF1 的核糖体结构也证明了 GGQ 位于肽转移酶中心的附近（图 14-37c）。目前尚不清楚 GGQ 是直接参与了多肽链从肽酰 tRNA 中水解，还是诱导了肽转移酶中心的某些变化，使得这一中心本身能催化水解。对肽酰转移酶中心（例如 A2541 或 A2602）靠近 CCA 末端所发现的保守碱基的研究表明，其中几个残基是肽的水解所必需的。实际上，较之在肽键形成中的作用，这些碱基似乎在肽的释放中具有更重要的作用。关于这种差别的一个可能的解释就是，只有邻近的 RNA 残基可以摆放一个小的水分子用于水解，但核糖体上的许多位点上的残基能够帮助摆放较大的 tRNA 分子用于催化。

总之，这些研究提出了以下假设：Ⅰ类释放因子是在功能上模仿了 tRNA，它具有一个与终止密码子作用的肽反密码子以及能进入肽转移酶中心的 GGQ 系列。RF1 与 tRNA 的结构对比揭示了这一蛋白怎样从功能上模拟了一个 tRNA（图 14-38）。正如 CCA 的 3′端和反密码子环占据每个 tRNA 的最末端那样，GGQ 和肽“反密码子”环也占据 RF1 的最末端。

GDP/GTP 交换和 GTP 水解调控Ⅱ类释放因子的功能

一旦Ⅰ类释放因子启动了肽酰 tRNA 的水解，它必须离开核糖体（图 14-39）。这一步是由Ⅱ类释放因子 RF3（或在真核细胞中 eRF3）来实现的。RF3 是一种 GTP 结合蛋白，但与其他参与翻译的 GTP 结合蛋白不同，这一因子与 GDP 的亲和力大于与 GTP 的亲和力。因此，游离的 RF3 主要是以与 GDP 结合的形式存在。RF3-GDP 与核糖体结合取决于Ⅰ类释放因子存在与否。在Ⅰ类 RF 激发多肽释放后，核糖体和Ⅰ类因子的构象改变诱导 RF3 交换 GDP 而结合 GTP。也就是说，这些因子扮演 RF3 的 GTP 交换因子，就像 EF-Ts 扮演 EF-Tu 的 GTP 交换因子那样。RF3 与 GTP 的结合导致其

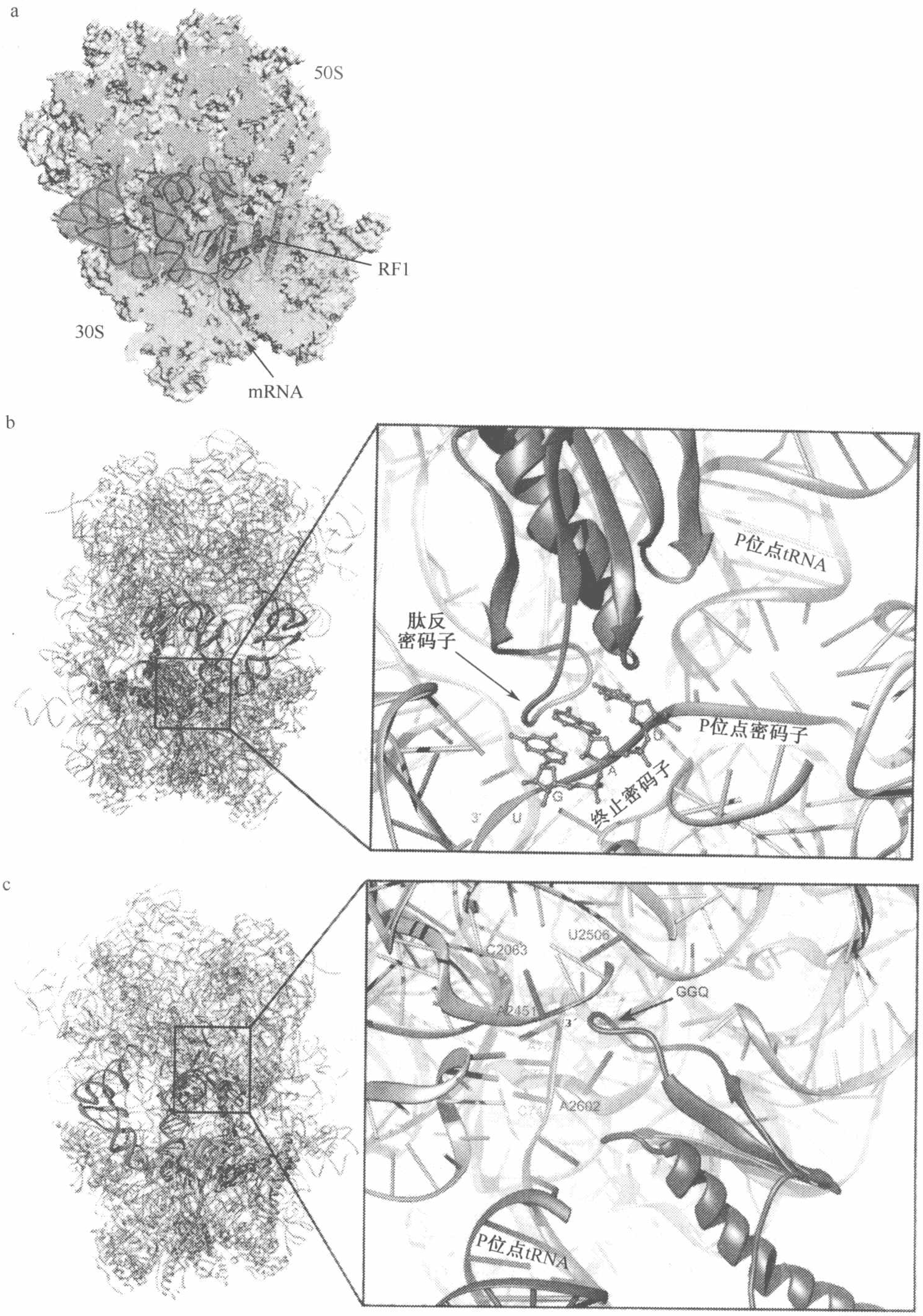


图 14-37　结合 RF1 的核糖体三维结构。 a. 本图显示结合于核糖体 A 位点的 RF1。b. 这一结构显示肽反密码子与反密码子非常接近。c. 本图显示，RF1 与核糖体结合的结构中，GGQ 基序与 P 位点 tRNA 的 3′端和肽酰转移酶中心十分接近。（来源：Petry et al. 2005. *Cell* 123：1255-1266. © Elsevier.）

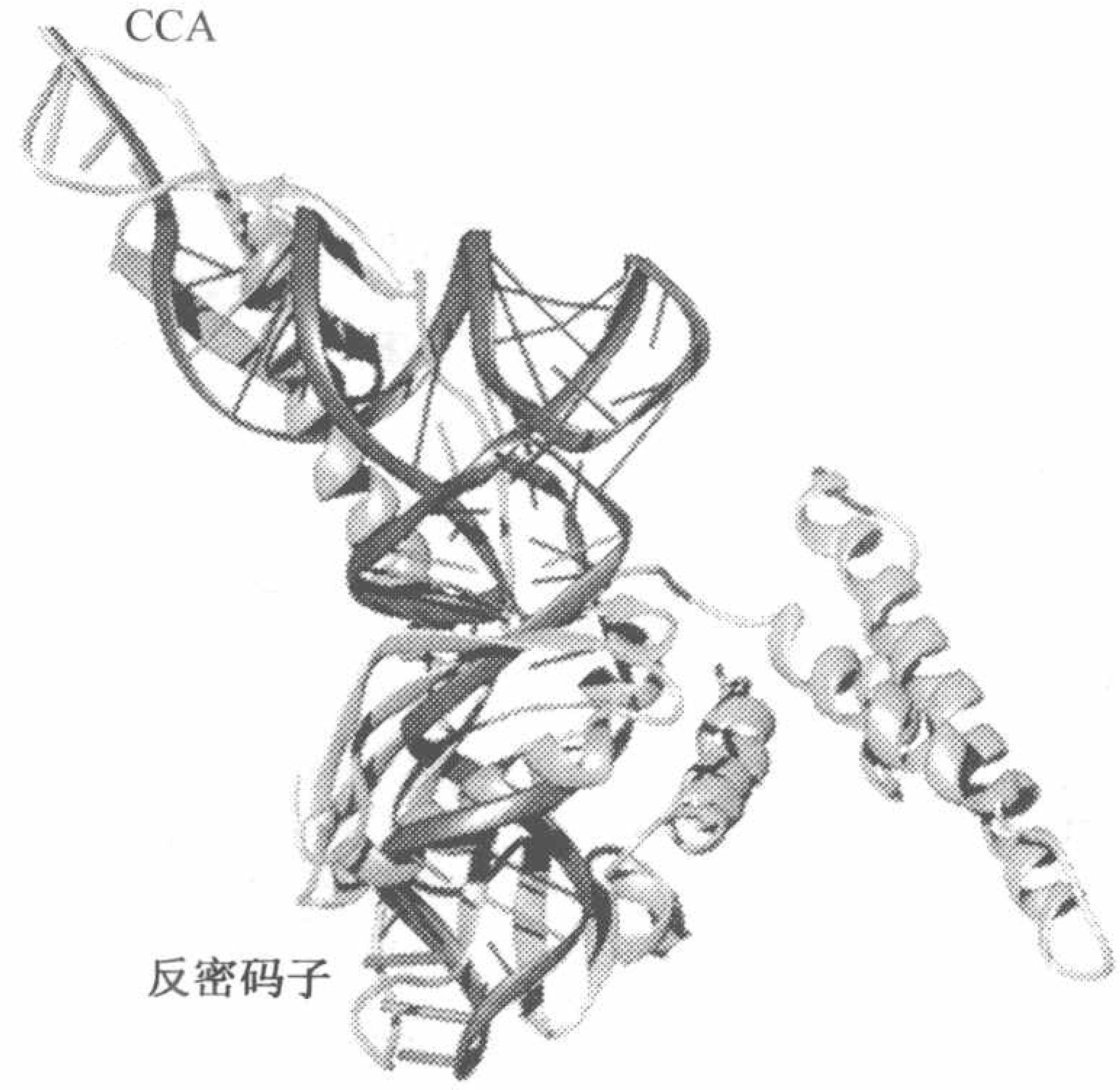


图 14-38 RF1 与 tRNA 的结构比较。tRNA 以深红色表示，RF1 占据同一位置，以灰色表示。（经同意，由 Petry et al. 2005. *Cell* 123：1255-1266，图 3E 重新绘制而成。© Elsevier.）

与核糖体之间的高亲和力相互作用的形成，从核糖体上并替换Ⅰ类分子。这一变化还使得 RF3 与大亚基的因子结合中心相结合。与参与翻译的其他 GTP 结合蛋白一样，这个作用引发 GTP 的水解。在（与核糖体结合的）Ⅰ类因子不存在的情况下，RF3-GDP 与核糖体的亲和力较弱而很快从核糖体中释放出来。

核糖体回收因子模仿 tRNA

在多肽链和释放因子离开核糖体后，核糖体仍然结合 mRNA，并留有两个脱氨基的 tRNA（在 P 位点和 E 位点）。为了参与新一轮的多肽合成，tRNA 和 mRNA 必须离开核糖体而核糖体也必须分解为大、小亚基。这几个事件统称为核糖体循环（ribosome recycling）。

在原核细胞中，一种称为核糖体循环因子（ribosome recycling factor，RRF）的蛋白质在多肽释放后与 EF-G 和 IF3 共同作用来使核糖体循环（图 14-40）。RRF 与空位的 A 位点结合，并模仿 tRNA。RRF 也募集 EF-G 到核糖体上，在那些模仿延伸过程中 EF-G 功能的事件中，EF-G 引发结合在 P 位点和 E 位点的空载 tRNA 的释放。虽然这一释放过程的详细机制并不清楚，但一般认为 A 位点的 RRF 被 EF-G 所替代的方式就像延伸过程中 A 位点的 tRNA 被 EF-G 所替代一样。一旦 tRNA 从核糖体中脱离，EF-G-GDP 和 RRF 同 mRNA 一起也从核糖体中释放出来。IF3（起始因子）也可能参与了 mRNA 的释放，而且它对核糖体分解为大、小亚基是必需的。这些事件的最后结果是形成一个与 IF3（而非 tRNA 或 mRNA）结合的小亚基和一个游离的大亚基。这些游离的核糖体就可以参与新一轮的翻译。

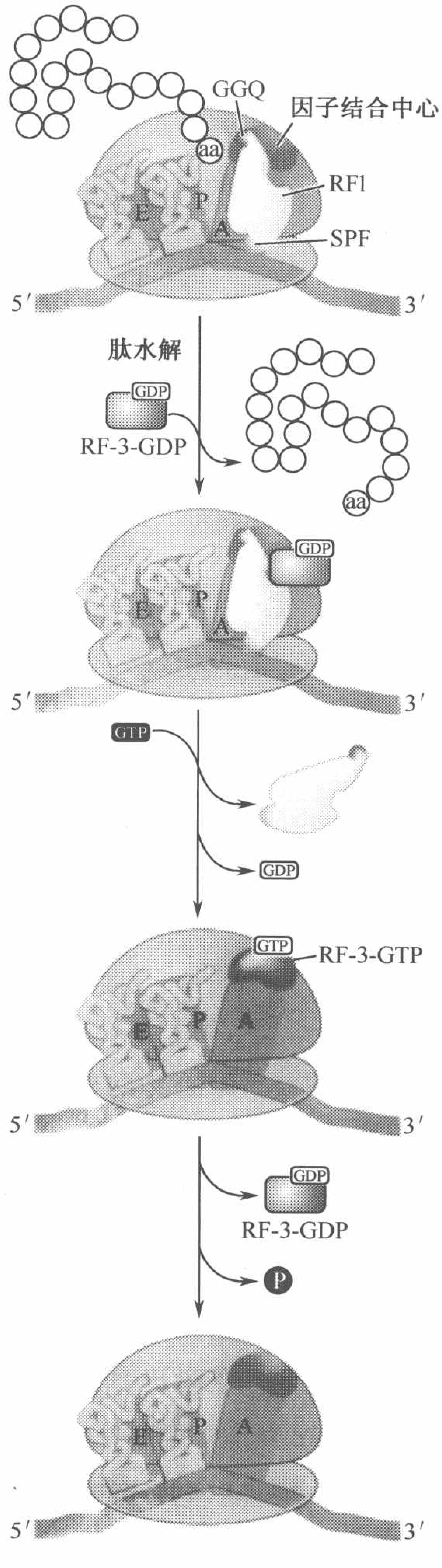


图 14-39 两个释放因子催化多肽的释放。Ⅰ类释放因子（RF1）识别终止密码子，并激发多肽通过位于肽酰转移酶中心的 GGQ 基序释放。Ⅱ类释放因子（RF3）只在多肽释放后与核糖体结合，并驱使Ⅰ类释放因子的解离。

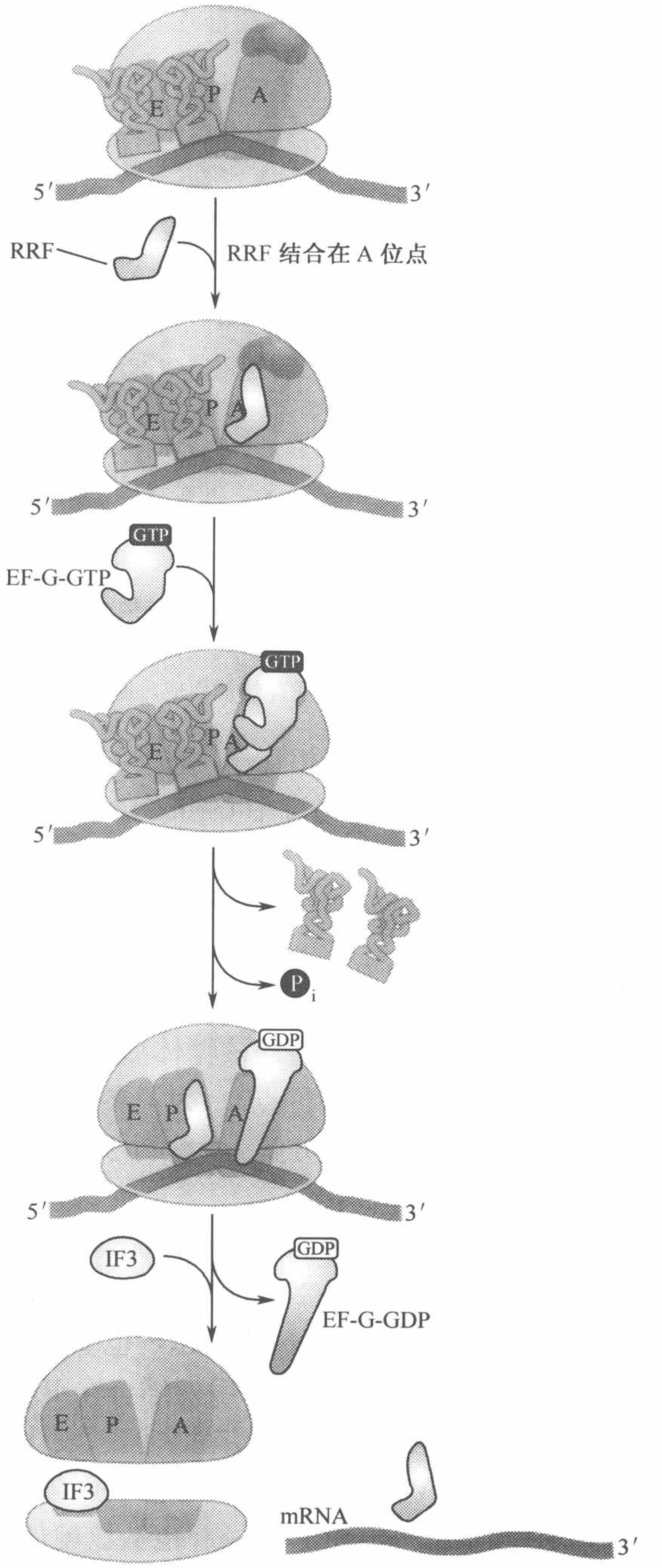


图 14-40 RRF 和 EF-G 联合激发 tRNA 和 mRNA 从终止的核糖体中释放出来。

重新强化一下 RRF 是 tRNA 模仿者的观点，事实上 RRF 的三维结构与 tRNA 相似。但是其与核糖体作用的方式与 tRNA（与核糖体）作用的方式是非常不同的。RRF 仅与 A 位点的大亚基部分紧密结合。我们可以对循环因子和 tRNA 作用的进行如下分析：如果核糖体循环因子完全模仿 A 位点的 tRNA，那么 P 位点的 tRNA 将被 EF-G 移至 E 位点。相反，EF-G 和循环因子将 P 位点的 tRNA 直接从 P 位点释放出来。很有可能，EF-G 和循环因子引起了核糖体结构的变化比一般易位过程所发生的要大，导致 mRNA 和 tRNA 同时释放出来。

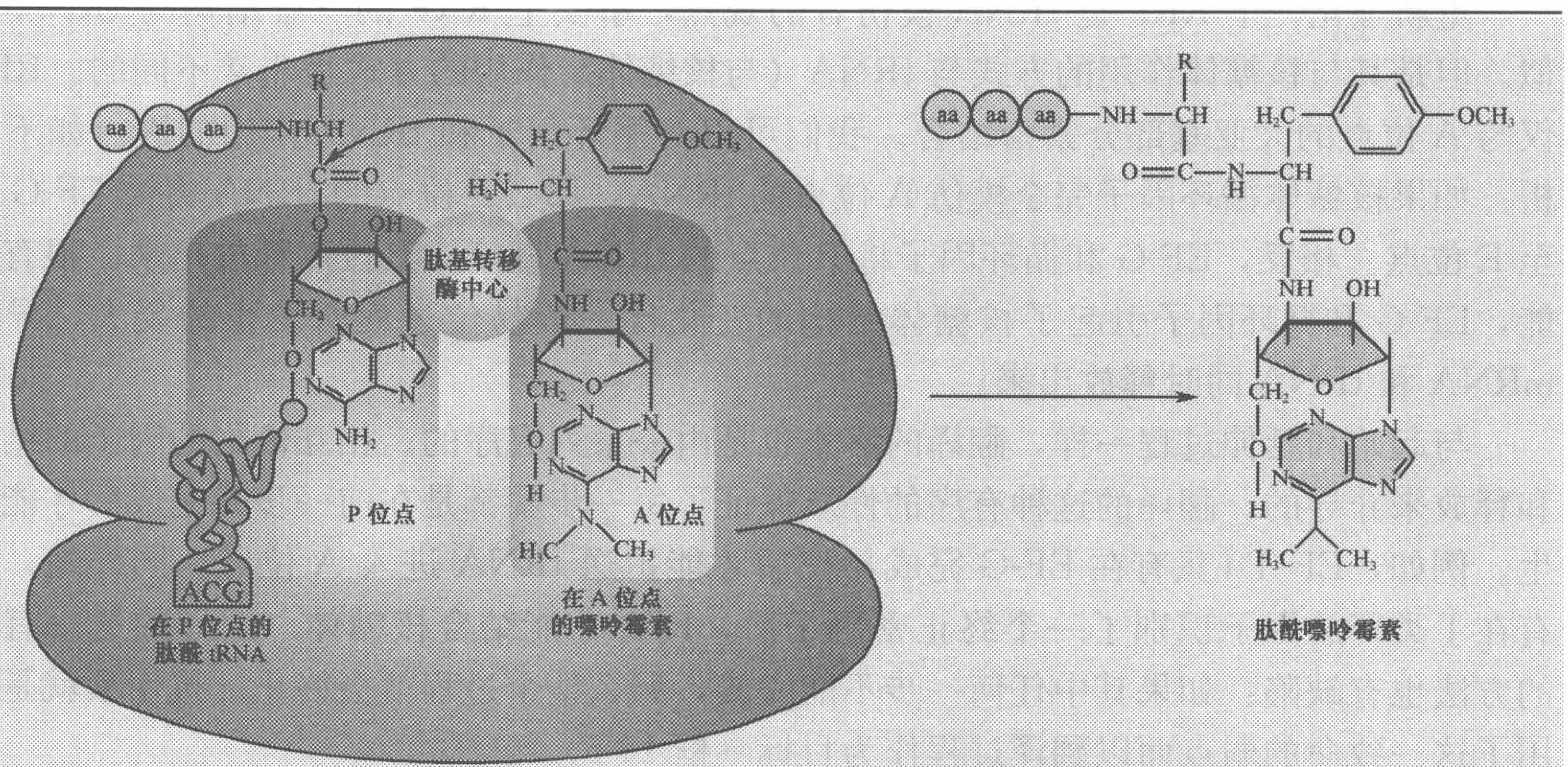
与起始和延伸过程一样，翻译的终止也是由一系列有序的、相互依附的因子的结合和释放来介导的。翻译的这种有序的性质保证了每一步骤都是在前一步骤完成后才能发生。例如，EF-Tu 只有在 EF-G 完成易位后才能护送 tRNA 进入 A 位点。同样地，只有在Ⅰ类释放因子识别了一个终止密码子后，RF3 才能结合核糖体。翻译的这一有序的方法也有缺陷：如果其中任何一步不能完成，那么整个过程就会终止。抗生素就是利用了这一致命的弱点而以翻译过程作为目标（框 14-5）。

框 14-5　抗生素通过阻遏翻译的某些特定步骤抑制细胞分裂

抗生素代表了与疾病搏斗的一种有力工具。在医学中最为广泛使用的抗生素杀死细菌但对真核细胞没有或很少有作用，因而对患者是无毒性的。自从 20 世纪上半叶抗生素发现以来，已经使诸如肺结核、细菌性肺炎、梅毒和淋病等以前无法治疗的感染变得可治疗了（尽管抗生素抗性细菌的出现正在成为有效治疗越来越大的障碍）。抗生素在细菌细胞里有不同的目标，但大约 40%的已知抗生素是翻译机器的抑制因子（框 14-5 表 1）。一般来讲，抗生素结合翻译机器的一个组分并抑制它的功能。因为不同的抗生素在不同的步骤中阻遏翻译，并且非常精确地进行（例如，就在 EF-Tu 释放之前），这些药剂已经成为研究蛋白质合成机制的有力工具。这样，除了它们明显的医学好处外，抗生素已经在帮助我们理解翻译机器的工作原理上扮演了一个重要的角色。

嘌呤霉素是一个经常用在翻译研究中的抗生素。它结合在 A 位点的大亚基区域。一旦结合，嘌呤霉素可以在肽基转移反应中取代氨基酰 tRNA（框 14-5 图 1）。因为嘌呤霉素相对于 tRNA 来说非常小，它结合到 A 位点不足以把多肽链保留在核糖体上。于是，转移到嘌呤霉素上的肽链以不完整的、结合嘌呤霉素的形式从核糖体上分离下来。换句话说，嘌呤霉素导致多肽合成未成熟时即终止了。其他抗生素以核糖体的其他特征作为目标，如肽出口通道、肽基转移中心、因子结合中心、解码中心和对易位重要的区域。

还有些抗生素是翻译因子的抑制因子。例如，黄色霉素和梭链孢酸分别是延伸因子 EF-Tu 和 EF-G 的抑制因子（框 14-5 表 1）。在两种情况中，抗生素都与结合 GTP 的翻译因子相互作用，阻止通常在 GTP 水解后发生的构象改变。于是，黄色霉素使核糖体停留在与 EF-Tu · GDP 氨基酰 tRNA 结合的形式。同样地，梭链孢酸使核糖体停留在与 EF-G · GDP 结合的形式。在两种情况中，翻译的下一步都因为延伸因子无法释放而被阻挠了。



框 14-5 图 1　嘌呤霉素通过在 A 位点模仿 tRNA 来终止翻译。嘌呤霉素结合在 A 位点上并参与肽键形成。一旦完成（肽键形成），嘌呤霉素和相连的多肽就从核糖体上散发出来。

框 14-5 表 1　抗生素：目标和后果

抗生素/毒素	目标细胞	分子目标	后果
四环素	原核细胞	30S 亚基 A 位点	抑制氨基酰 tRNA 结合到 A 位点上
潮霉素	原核和真核细胞	30S 亚基 A 位点附近	阻挠 A 位点 tRNA 到 P 位点的易位
巴龙霉素	原核细胞	靠近 30S 亚基 A 位点的密码子-反密码子相互作用位点	通过降低密码子-反密码子配对的敏感度在翻译过程中提高错误率
氯霉素	原核细胞	50S 亚基的肽基转移酶中心	阻断 A 位点的氨基酰 tRNA 为进行肽基转移反应的正确定位
嘌呤霉素	原核和真核细胞	核糖体大亚基的肽基转移酶中心	链终止子；模仿 A 位点的氨基酰 tRNA 的 3′端，作为新的多肽链的接受者
红霉素	原核细胞	50S 亚基的肽出口通道	关闭成长中的多肽链从核糖体离开的出口；阻遏翻译
梭链孢酸	原核细胞	EF-G	阻挠 EF-G-GDP 从核糖体中释放
硫链丝菌肽	原核细胞	50S 亚基的因子结合中心	妨碍 IF2 和 EF-G 与因子结合中心的接合
黄色霉素		EF-Tu	阻挠与 GTP 水解关联的构象改变，从而防止 EF-Tu 的释放
蓖麻毒素和 α-帚曲菌素（蛋白质毒素）	原核细胞和真核细胞	在核糖体大亚基的因子结合中心化学修饰 RNA	阻挠翻译因子 GTP 酶的活化
白喉毒素	真核细胞	化学修饰 EF-Tu	抑制 EF-Tu 的功能
放线（菌）酮	真核细胞	60S 亚基的肽基转移酶中心	抑制肽基转移酶活性

翻译的调控

尽管大多数基因的表达都是在mRNA转录水平上进行调控的，但有些情况下对细胞来说从蛋白质合成水平上来进行调控更有效。翻译调控较之转录调控的优点之一是能够对外界刺激迅速产生反应。蛋白质合成水平的调控免去了改变mRNA转录水平所需的时间（而且在真核生物中还除去包括mRNA加工及运输到胞质的时间），因此使蛋白水平的改变更加迅速。像其他类型的调控那样，翻译调控主要在起始步骤。总体来说，调控一个通路的早期步骤较之起始后再停止反应更加有效。对翻译来说，起始水平的调控也排除产生那些改变了功能的不完整蛋白产物。

在本部分中，我们将首先描述细菌和真核细胞翻译调控所使用的一般机制。而后，我们介绍使用这种调控的一些具体实例。

蛋白质或RNA在核糖体结合位点附近结合下调细菌翻译的起始

细菌翻译起始调控因子的首要目标就是干扰30S亚基对RBS的识别。总体来说，这些抑制因子的作用机制就是与RBS附近的序列结合，从物理上阻断了RBS和16S rRNA的碱基配对（图14-41a）。这些抑制因子通常是RNA结合蛋白，识别形成于RBS附近的RNA结构。尽管它们不直接与RBS结合，但结合蛋白足够大从而在空间

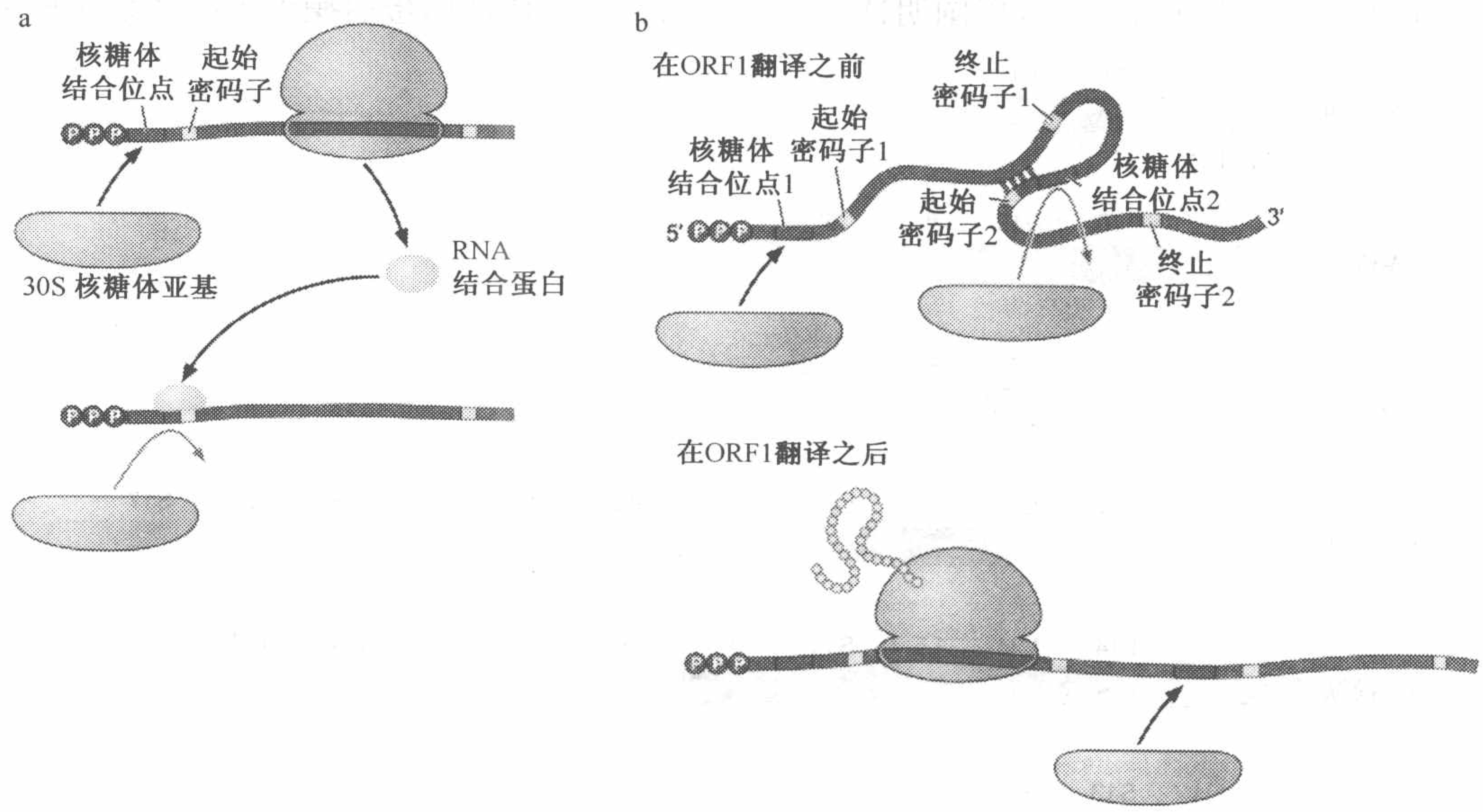


图14-41 通过抑制30S亚基的结合来调控细菌翻译起始。 a. 结合于RBS附近结点的蛋白阻止16S rRNA接近RBS。在这种情况下，mRNA编码的蛋白结合于自身mRNA的RBS。b. mRNA分子内的碱基配对能够干扰与16S rRNA的碱基配对。在很多情况下，这种抑制由处于同一操纵子内的其他基因的翻译来调节。如果与RBS邻近区域碱基配对的mRNA区域处于某个ORF内，那么当这个ORF翻译时，起干扰作用的碱基配对发生断裂，使得下一个核糖体能够识别先前被掩盖的RBS。

阻止 30S 亚基接近 RBS。实际上，这些抑制因子不直接与 RBS 结合有着重要意义，因为如果那样就会产生阻止细胞中大部分蛋白翻译的风险。

RNA 分子也可以利用相同机制来充当翻译的抑制因子。这种调控经常发生在 mRNA 与自己的碱基配对，掩盖了一个或更多的 RBS（图 14-41b）。这种掩盖可以抑制相关 ORF 的翻译，直到这种相互作用被中断。在很多情况下，中断的发生是操纵子翻译另一个基因导致的后果。在这种情况下，与 RBS 邻近区域相互作用的 mRNA 区域存在于另一个 ORF 中，核糖体的通过阻断这种碱基配对，因此使得另一个核糖体可以识别未被掩盖的 RBS。

原核翻译调控：核糖体蛋白是其自身合成的翻译抑制因子

现住我们介绍一个细菌中翻译调控的例子，阐明细胞怎样利用这一机制控制核糖体蛋白基因的正确表达。协调核糖体蛋白的表达与 rRNA 的表达向细胞提出了一个有趣的调控问题。正如我们前面谈到的，每个核糖体都包含 50 多种截然不同的蛋白，它们应该和与之结合的 rRNA 以相同速率合成。此外，细胞合成蛋白的速率及其需要的核糖体数目与细胞的生长速率紧密相关。生长环境的改变会迅速导致所有核糖体组分合成速率的适当升高或降低。而这种协调调控是怎样完成的呢?

核糖体蛋白基因的协调调控通过组成若干操纵子而得以简化，每个操纵子包括最多 11 个核糖体蛋白基因（图 14-42）。与其他操纵子一样，这些基因簇是在 RNA、合成水平进行调控的（第 16 章）：然而，最重要的核糖体蛋白合成调控是在 mRNA 翻译水平进行的。一个简单的实验可以阐明这个结论。当把一个核糖体蛋白操纵子的多余拷贝引

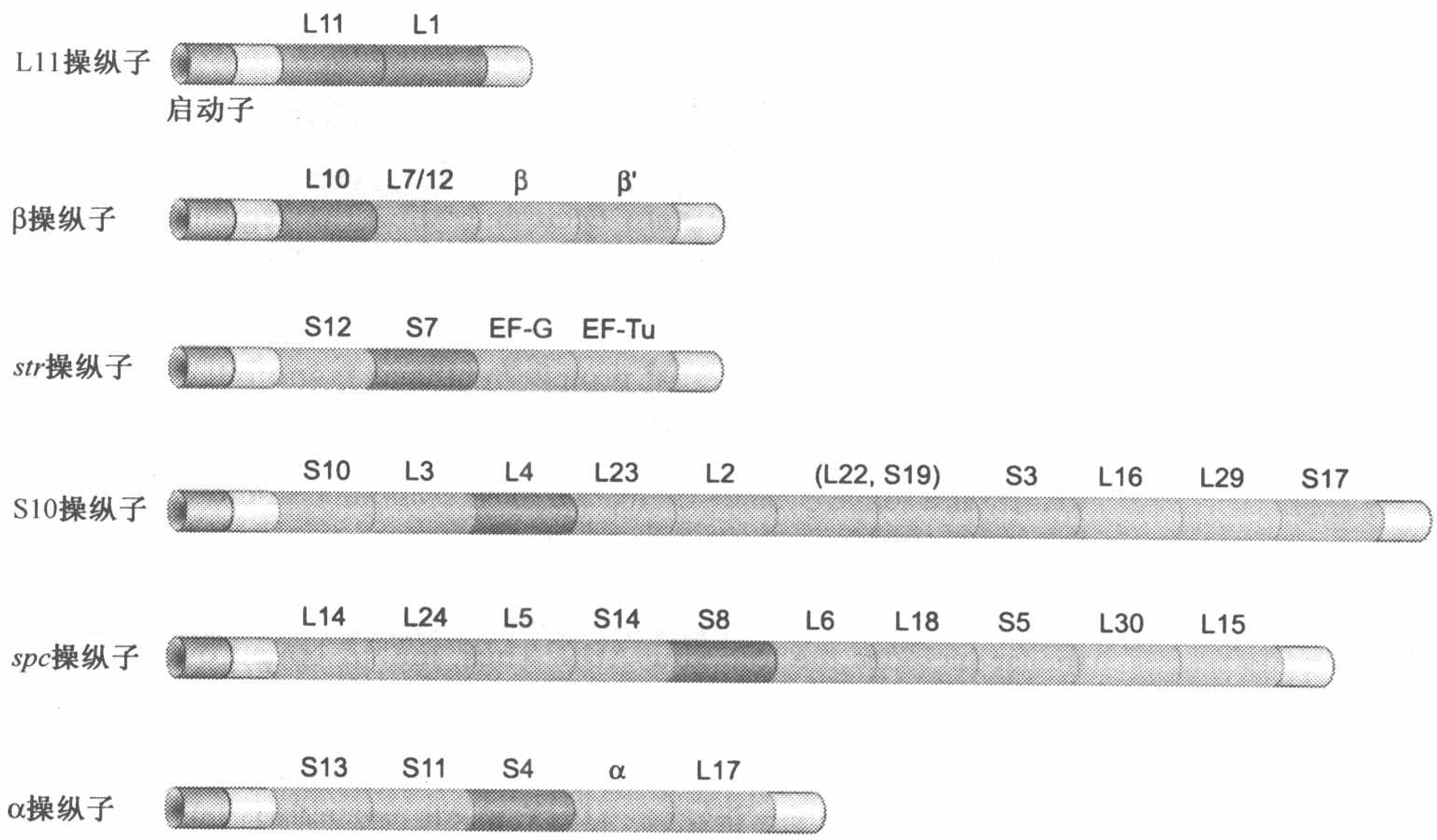


图 14-42 *E. coli* 核糖体蛋白操纵子。作为其他蛋白的翻译抑制因子的蛋白以红色阴影表示。启动子以紫色表示。每个操纵子都按照编码的核糖体蛋白来标记（比如，L14 是较大核糖体蛋白 14）。（征得同意，采自 Nomura M. et al，1984. *Annu. Rev. Biochem.* 53：75-177. © Annlual Reviews.）

进细胞时，其 mRNA 的数目相应增多，但核糖体蛋白的合成几乎没变。这样，细胞通过减少 mRNA 作为蛋白合成模板的使用来协调平衡多余的 mRNA。

核糖体蛋白 mRNA 翻译的严格调控是自我抑制的结果。对于每个核糖体蛋白操纵子来说，一个编码的核糖体蛋白（或两个编码的核糖体蛋白的复合体）在最接近 5′基因的其中一个翻译起始序列附近结合那一操纵子的 mRNA。核糖体蛋白的结合在空间上抑制了核糖体小亚基与附近 RBS 的结合，因此也就抑制了翻译的起始。

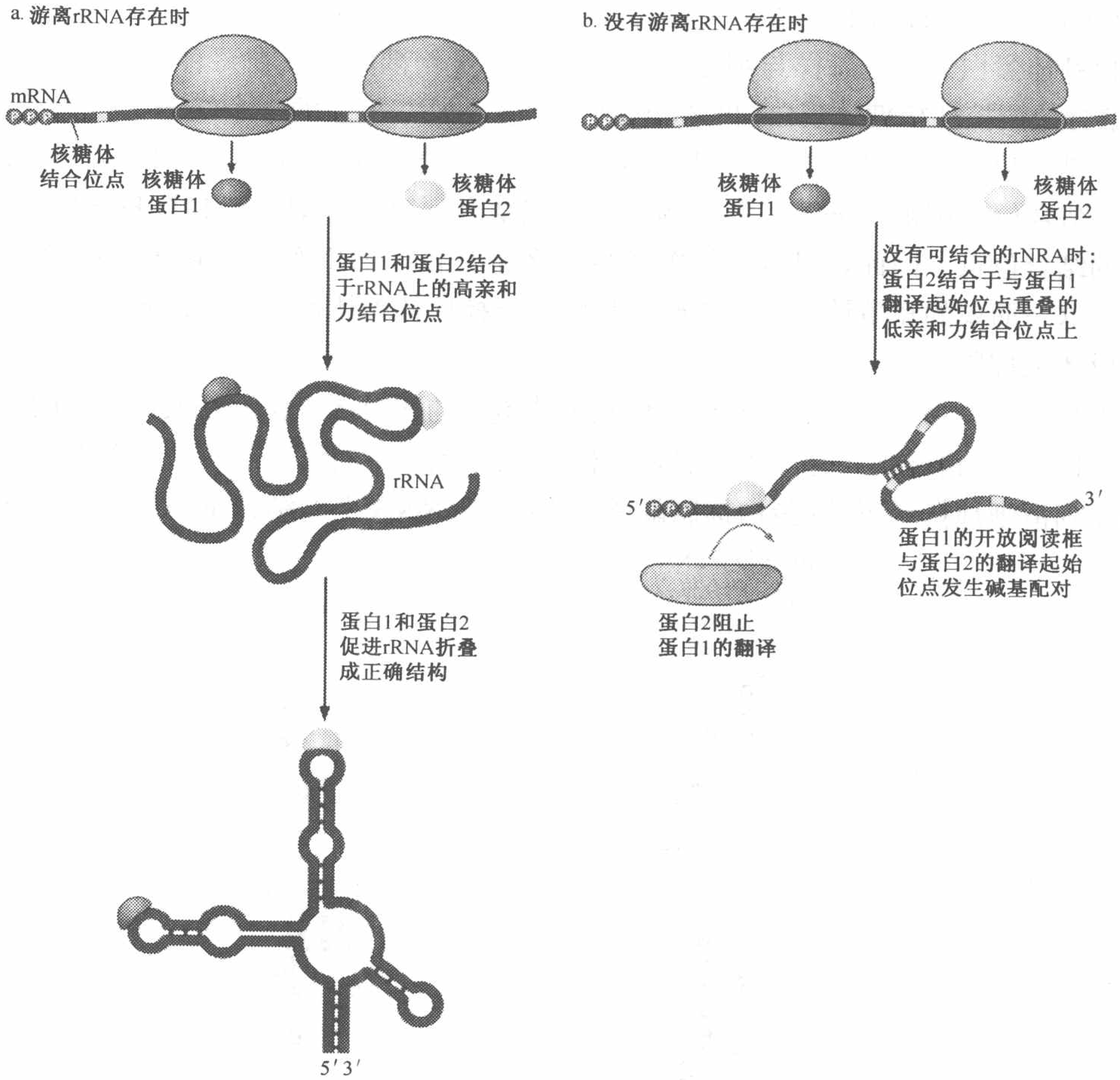


图 14-43　核糖体蛋白表达的调控。在这个例子中，我们介绍一个简化的双蛋白核糖体蛋白基因操纵子。a. 在没有结合的 RNA 存在的情况下，这个操纵子表达的两个核糖体蛋白都与 rRNA 结合，并且协助核糖体的正确装配。在这种情况下，核糖体蛋白 2 不与邻近核糖体蛋白 1 RBS 的较低亲和力结合位点结合。b. 在没有 rRNA 让核糖体蛋白结合时，核糖体蛋白 2 就结合于邻近的核糖体蛋白 1 的 RBS，从而阻断这一 ORF 的翻译。在没有翻译核糖体蛋白 1 存在时，与核糖体蛋白 2 RBS 邻近序列互补的序列就与之发生碱基配对，阻止 30S 亚基与此位点的结合。如正文所述，翻译耦联也可发挥抑制下游 ORF 翻译的作用。

理解核糖体蛋白的结合怎样阻止操纵子起始基因的翻译是十分容易的。但这种结合是怎样影响下游那些拥有其自身 RBS 的基因呢？这种“极性”效应可以通过多种机制发生。

正如我们在本章前面部分所介绍的，当一个上游基因的终止密码子距离下游基因的起始密码子很近时，就可能发生耦联翻译。这种邻近性所产生的情况是下游基因的翻译需要上游基因的翻译才能得以完成。第二种机制则是利用 mRNA 折叠成一定的结构。核糖体蛋白操纵子 mRNA 通常折叠成这样的结构，即只有当 mRNA 的前端基因被翻译，才允许识别内部的 RBS。例如，假定 mRNA 第一个基因编码区的一个区域与第二个基因的 RBS 附近的一个位点发生碱基配对。在这种情况下，只有当抑制性的碱基配对因核糖体住第一个基因的编码区进行翻译而中断之后，16SrRNA 才识别第二个基因的 RBS（图14-43）。

核糖体蛋白的表达怎样与细胞中 rRNA 的数量相耦联呢？在每种情况下，与 mRNA 结合的调控性核糖体蛋白同时也识别位于适当 rRNA 上的一个结合力很强的位点。若这一结合位点没有被占据，核糖体蛋白会首先结合这个位点。另一方面，若所有的 rRNA 结合位点都被占据了，那么调节蛋白就结合到第二个即位于其自身 mRNA 上的较弱的结合位点。因此，只有当核糖体蛋白多于其目标 rRNA 时，它才与自身 mRNA 结合。这一简单的竞争性结合事件作确保只有当调控性核糖体蛋白数量过剩时核糖体蛋白的合成才被抑制。

不足为奇，在很多情况下，调控性核糖体蛋白的两个结合位点是彼此相关的。在核糖体蛋白 S8 的情况中，这两个位点具有相当大的相似性（图 14-44）。mRNA 上结合位点的序列清晰地揭示了 S8 蛋白抑制翻译的机制。mRNA 上的结合位点包括起始 AUG。

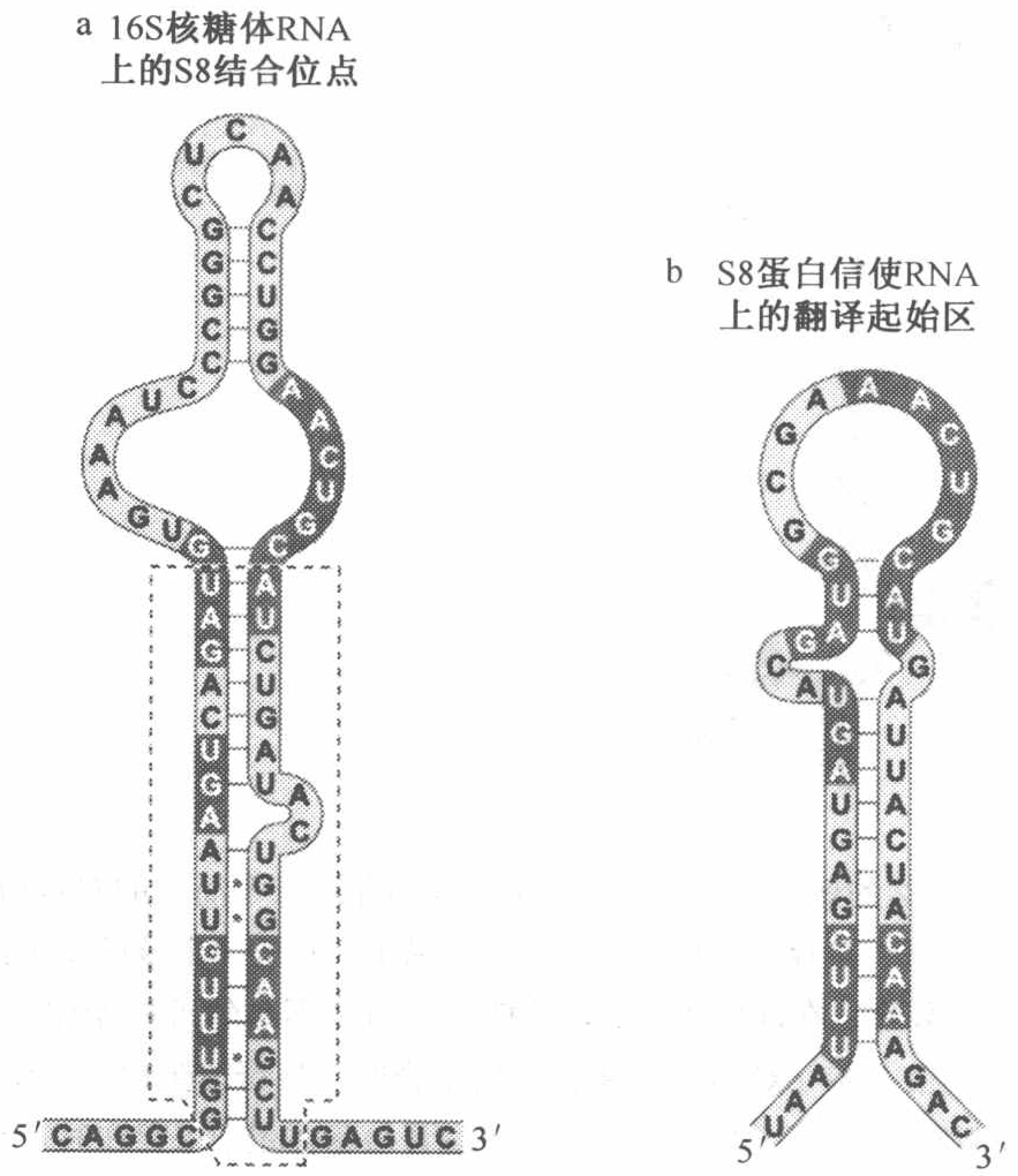


图 14-44 核糖体蛋白 S8 与 16S rRNA 及其自身 mRNA 结合。此图为核糖体蛋白 S8（由 *spc* 操纵子编码；图 14-43）与核糖体中 16S rRNA 结合的区域和其自身 mRNA 的翻译起始位点的对比。相似序列用深绿色表示。虚线框中表示的是 16S rRNA 被 S8 蛋白保护的区域。（来源：from Cerretti D. P. et al. 1988. *J. Mol. Biol.* 204：309-329. © Elsevier.）

因此，被过量的S8蛋白（在本例中）结合的mRNA不能结合核糖体而起始翻译。这两个结合位点的不同解释了为何与rRNA结合会较之与mRNA结合更紧密。因此，只有满足了核糖体装配对S8蛋白的需求后翻译才被抑制。

这种翻译调控的策略并不局限于核糖体蛋白。其他RNA结合蛋白通过结合自身mRNA来调节其自身的表达，包括一些氨酰-tRNA合成酶。此外，也有一些例子，mRNA根据细胞环境（例如，温度或代谢水平，详见第18章）折叠成不同的结构或促进翻译，或抑制翻译。

真核翻译总体调节因子以mRNA识别及起始tRNA核糖体结合所需的主要因子为靶标

在营养物质不足或其他细胞应激情况下，对于真核细胞来说从总体上降低翻译水平通常是十分有用的。在这些情况下，真核翻译起始的两个早期步骤成为抑制的靶标：mRNA的识别以及起始tRNA与40S亚基的结合。前面说过，在真核翻译起始中，这两个事件彼此独立地发生，但抑制任何一个步骤均会抑制新蛋白的合成。在两种情况下，抑制的机制都是由磷酸化控制的。

一个常见的抑制机制是由eIF2的磷酸化介导的。前面讲过，eIF2结合于GTP是将起始tRNA运输到真核核糖体40S亚基的P位点所必需的。很多使eIF2的α亚基磷酸化的蛋白激酶已经被鉴定出来。这一亚基的磷酸化抑制了eIF2的一个叫做eIF2B的GTP交换因子的作用，导致eIF2-GTP水平的降低。与EF-Ts对EF-Tu-GDP的作用相似，eIF2B激活eIF2-GDP以释放结合的GDP并与GTP结合。因为eIF2与GTP的结合是护送起始tRNA到40S亚基所必需的，所以eIF2-GTP水平的降低限制了翻译的起始。已知的eIF2 α激酶可被众多不同的胞内环境所激活，包括氨基酸缺乏（见下）、病毒感染及温度升高。

另一种常见的翻译抑制机制以5′帽结合蛋白eIF4E为靶标。前面讲过，在与5′帽结合之后，eIF4E结合eIF4G。eIF4G中识别eIF4E的短结构域在一个叫做eIF4E结合蛋白或4E-BPs的蛋白家族中也有发现。这些蛋白与eIF4G竞争结合eIF4E并因此充当翻译的一般抑制因子（图14-45）。与eIF2类似，4E-BPs也受磷酸化的调控。在其非磷酸化状态下，4E-BPs与eIF4E紧密结合，抑制翻译的进行。相反，4E-BPs的磷酸化抑制其与eIF4E的结合（图14-45）。

4E-BP的磷酸化由一个称为mTOr的关键的胞内蛋白激酶所介导。激活细胞分裂的生长因子、激素以及其他因子激活这种激酶，从而提高细胞的整体翻译水平。这些观察引发分裂一个假设，即翻译水平的调控与细胞的增殖是精确协调的。实际上，eIF4E的过表达可以导致细胞的癌性转化，并且mTor的抑制因子（如雷帕霉素）是十分有效的化疗试剂。尽管我们是在翻译的总体调控的范畴中介绍的这些调控机制，但这两种机制也都用于细胞中特定mRNA的翻译调控，这将在下面叙述。

通过对mRNA特异的4E-BPs进行的翻译空间调控

除了翻译的总体调控，与eIF4E结合也用来调控特定mRNA的翻译。例如，黑腹

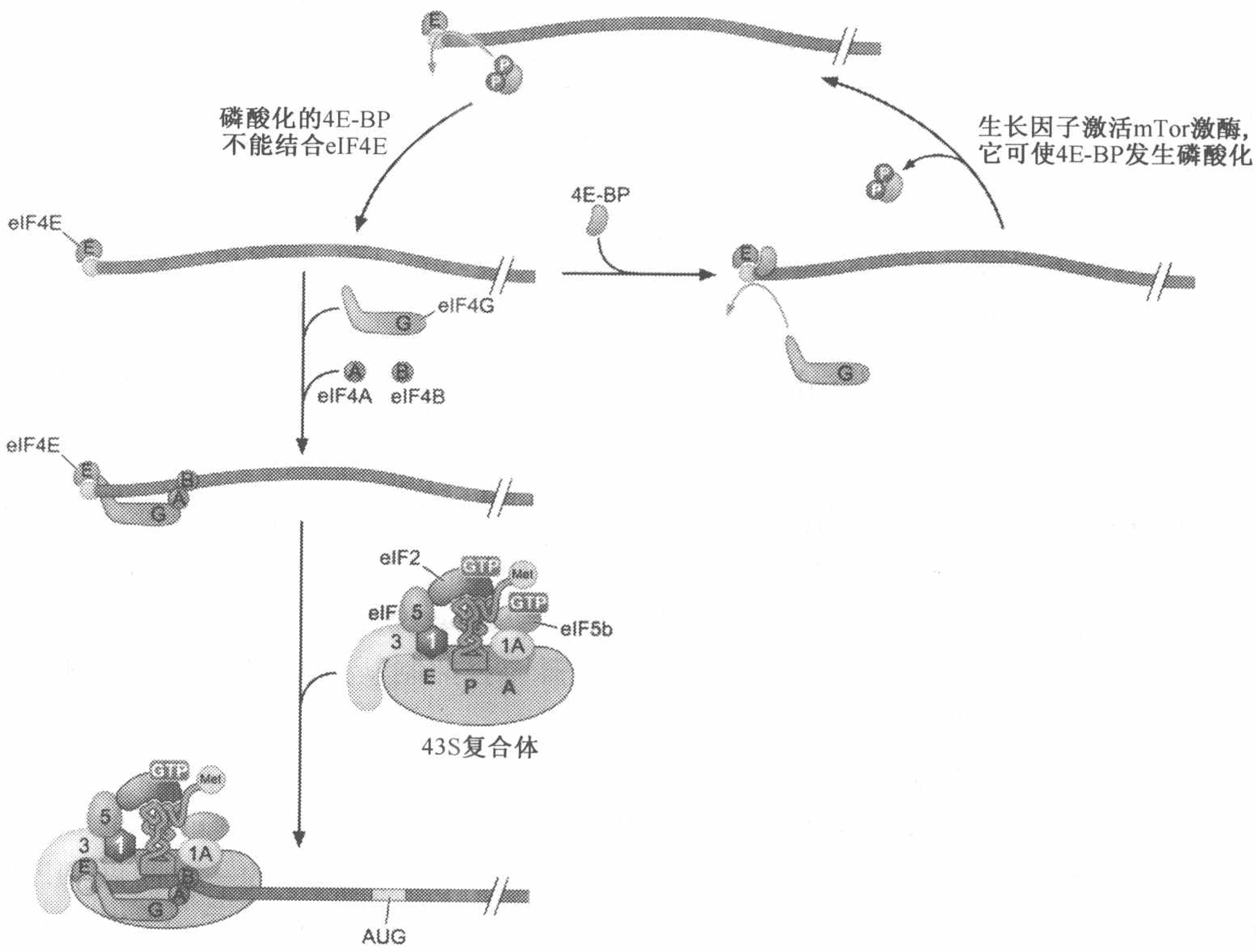


图 14-45 真核翻译起始总体上受到 eIF4E 结合蛋白（4E-BPs）的调控。 4E-BPs 与 eIF4G 竞争结合帽结合蛋白 eIF4E。这阻断了 eIF4A 介导的 mRNA 5′端的解旋及依赖的 eIF4G 43S 前起始复合物的募集。4E-BPs 与 eIF4E 的结合受磷酸化的调控。mTOr 激酶磷酸化 4E-BPs 的 eIF4E 结合区，使其再也不能识别 eIF4E。这样，通过抑制 4E-BPs 的作用，mTOr 激酶上调细胞的翻译活性。

果蝇卵母细胞及发育胚胎的前后轴的正确建立需要很多蛋白正确定位在一个共有的大细胞质中（见第 19 章对此事件的完整描述）。在有些情况下，这些关键调控蛋白的空间限制性翻译在控制其定位方面发挥着主要作用。

Oskar 蛋白在受精之前精确定位于卵母细胞的后部区域。尽管如此，Oskar mRNA 是由与母果蝇卵巢结合的滋养细胞合成的，并在受精前定位于卵母细胞的前部区域。随后，Oskar mRNA 被运输至卵母细胞的后部。细胞为了使 Oskar 蛋白的表达局限在卵母细胞后部，Oskar mRNA 在从卵母细胞的前部转移到后部时不被翻译是十分必要的。

一个称之为 Cup 的 4E-BP 的作用对特异性限制 Oskar mRNA 的翻译非常关键（图 14-46）。OskarmRNA 在 3′非翻译区包含若干序列，与一个叫做 Bruno 的蛋白特异结合。而 Bruno 可与 Cup 结合，并募集这一 4E-BP 结合到 Oskar mRNA 上。当定位于 Oskar mRNA 后，Cup 竞争过 eIF4G 而结合到 eIF4E，因此抑制了 mRNA 的翻译。Cup 并没有足够多的数量像上述总体调控的 4E-BP 那样作用于所有的翻译。尽管如此，当定位于某一 mRNA 后，Cup 即成为一个非常有效的翻译抑制因子。这一机制并不局限于 Oskar 蛋白。果蝇中的 Nanos 蛋白也是通过将 Cup 募集于其 mRNA 来调控的。与

此相似，一个叫做 CPEB 的 mRNA 结合蛋白将一种称为 Maskin 的 4E-BP 募集至许多在脊椎动物卵母细胞发育中翻译被抑制了的 mRNA 上。

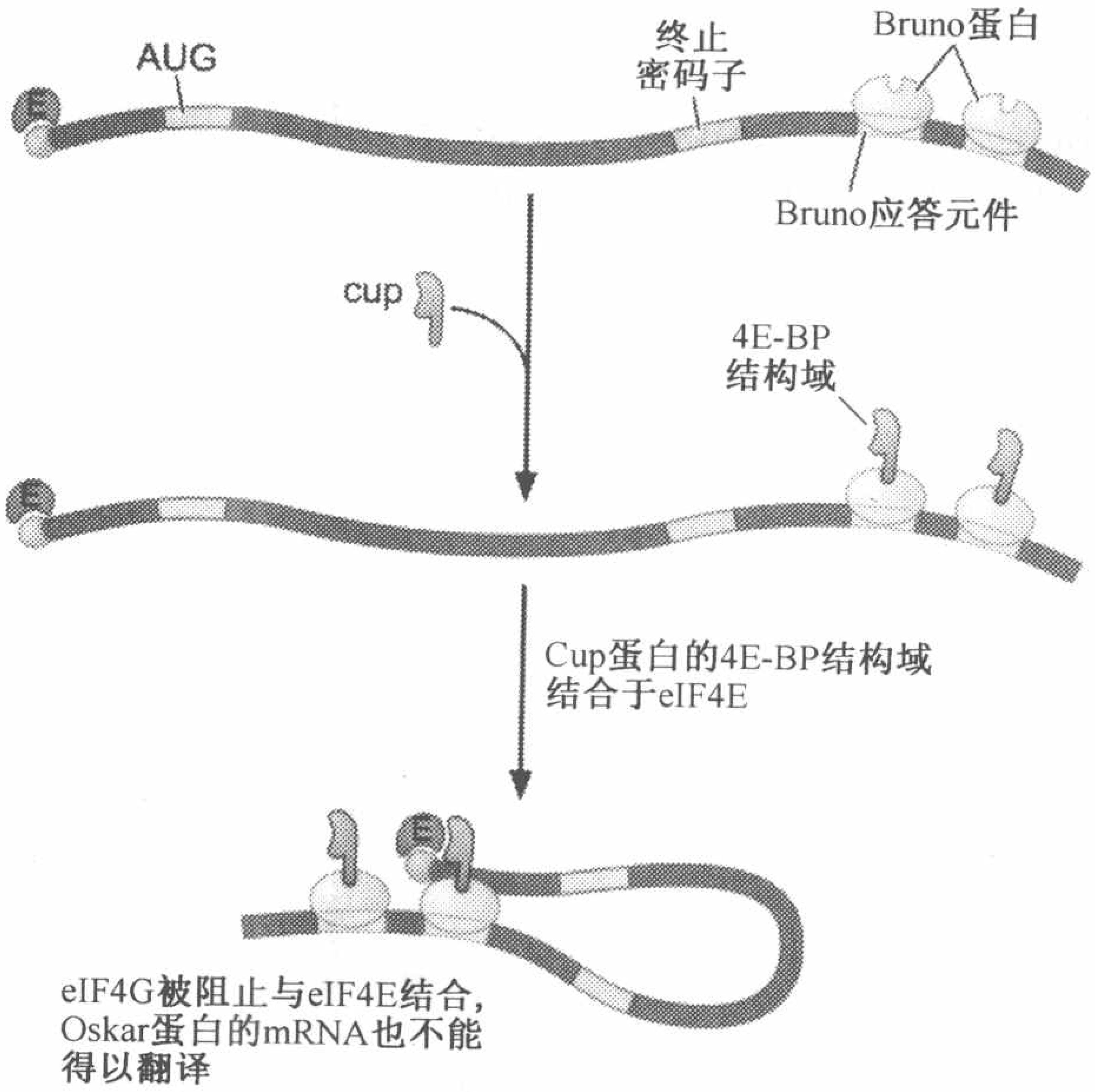


图 14-46 eIF4E 结合蛋白 Cup 特异性抑制 Oskar mRNA 的翻译。当 Oskar mRNA 从果蝇卵母细胞前部运输至后部的过程中，Oskar mRNA 不被翻译是十分关键的。Oskar 翻译的抑制由两个蛋白介导。RNA 结合蛋白 Bruno 结合于多个位于 Oskar mRNA 3′非翻译区的序列，它们称为 Bruno 应答元件（Bruno response elements，BREs）。随后 Bruno 募集 4E-BP Cup 到 mRNA。当定位于 mRNA 后，Cup 马上与 eIF4G 竞争结合 eIF4E，抑制这一 mRNA 的翻译。

一个铁离子调控的 RNA 结合蛋白控制转铁蛋白的翻译

调节人体内铁离子水平十分关键。许多蛋白利用铁离子作为辅助因子，包括血红蛋白和肌红蛋白等氧转运蛋白以及诸多参与氧化磷酸化的蛋白。与铁离子在氧气运输和能量生成中的重要作用相一致，人体中铁的缺乏（或称为贫血）会造成一种整体的虚弱感。而另一方面，过量的铁对细胞有毒害，并且可以导致肝损伤、心衰和糖尿病。因此，正确调节人体中铁离子水平是十分关键的。

铁离子结合蛋白转铁蛋白是人体中调节铁离子水平的主要调节因子。转铁蛋白以一种受控的方式贮存并释放铁离子，从而保持体内铁含量的稳定。因此，转铁蛋白的水平迅速应答于体内游离铁的水平很关键。对游离铁的水平变化作出迅速反应的要求使转铁蛋白表达的调节发生在蛋白合成水平。

转铁蛋白的翻译受一些称作铁离子调控蛋白（iron regulated proteins，IRP）的铁离子结合蛋白的调控。这些蛋白也是 RNA 结合蛋白，它们识别转铁蛋白 mRNA 的 5′端形成的一个叫做铁离子调控元件（iron regulatory element，IRE，图 14-47）的特异发卡结构。重要的是，这些蛋白识别 IRE 的能力受到细胞中铁离子水平的调控。在铁

离子缺乏的细胞中，铁离子的浓度太低而不能与 IRP 结合。在缺乏结合的铁离子的情况下，这些蛋白与 IRE 紧密结合，抑制 eIF4A/B 解开 IRE 发卡结构的能力。发卡结构的持续存在会在空间上成为 43S 复合体结合 mRNA 的阻碍。与此相反，当细胞中游离铁的浓度升高时，IRP 成为铁离子结合态。而结合了铁离子后，IRP 即失去其与 IRE 结合的能力并因此抑制翻译。总体来说，转铁蛋白 mRNA 的翻译水平受到 IRP 与游离铁离子亲和力及其对 IRP 结合于 IRE 的影响的严格调控。

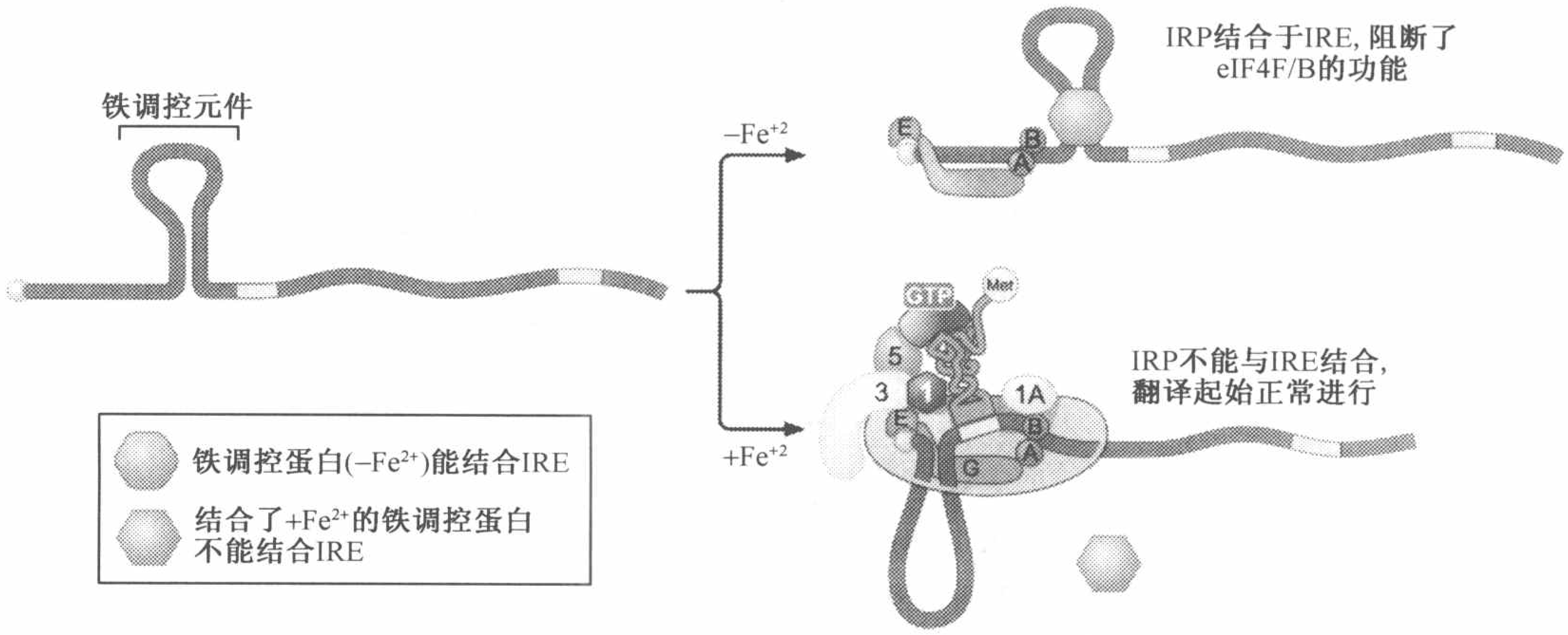


图 14-47 铁离子对转铁蛋白的翻译调控。 转铁蛋白基因 5′非翻译区包含一个叫做铁离子调控元件（IRE）的茎-环结构。当不与 Fe^{2+} 结合时，铁调控蛋白（IRP）就紧密结合于这一位点。通过稳定 IRE 的茎-环结构，IRP 阻止 eIF4A 从转铁蛋白 mRNA 末端清除这一结构。在这种情况下，43S 前起始复合体与 mRNA 的结合就不能发生，转铁蛋白基因也不被翻译。当铁离子水平升高并且需要转铁蛋白时，IRP 就与 Fe^{2+} 结合，这就抑制了 IRP 与 IRE 的结合能力，也因此使转铁蛋白得以翻译。

酵母转录激活因子 Gcn4 的翻译受到上游短 ORF 及三元复合物丰度的调控

Gcn4 是一个酵母转录激活因子，调控多种指导氨基酸生物合成的酶基因的表达。尽管 Gcn4 是一个转录激活因子，但其本身的调控却在翻译水平进行。在氨基酸水平低时，Gcn4 mRNA 得以翻译（所以生物合成的酶也得以表达）。但当氨基酸水平高时，Gcn4 mRNA 不被翻译。这样的调控是怎样做到的呢？

与典型的真核 mRNA 结构不同，编码 Gcn4 蛋白的 mRNA 在 Gcn4 编码序列上游有四个小的开放阅读框（称为 uORF）。其中最上游短 ORF（uORF1）可被从 mRNA5′端扫描的核糖体高效识别。一旦核糖体翻译了 uORF1，这一 ORF 的独有特性使得 50% 的核糖体小亚基仍然结合于 RNA，并且重新开始扫描下游以寻找起始密码子（AUG）（图 14-48）。

究竟哪个下游 AUG 被正在进行扫描的小亚基所识别，这受到进行扫描的小亚基何时结合于带有一个起始 tRNA 的 eIF2 复合物的调控。前面讲过，在 P 位点没有起始 tRNA 存在的情况下，40S 亚基就不能识别 mRNA 上的 AUG 序列。因此，在开始下游的任何一个 ORF 的翻译之前，进行扫描的 40S 亚基必须结合 eIF2 · Met-tRNA$_i^{Met}$（下面称为 TC，即 ternary complex 三元复合物）。

当氨基酸含量足够时，TC在核糖体翻译uORF1完成后不久就与正在进行扫描的核糖体重新结合（图14-18a）。一旦与TC重新结合，核糖体小亚基就可以识别一个AUG，在其他某一个uORF（2，3或4）上重新起始翻译。与uORF1的情形不同，在翻译完这些uORF之后，核糖体从mRNA上完全解离下来，失去翻译Gcn4 ORF的能力。因此，不产生Gcn4蛋白。

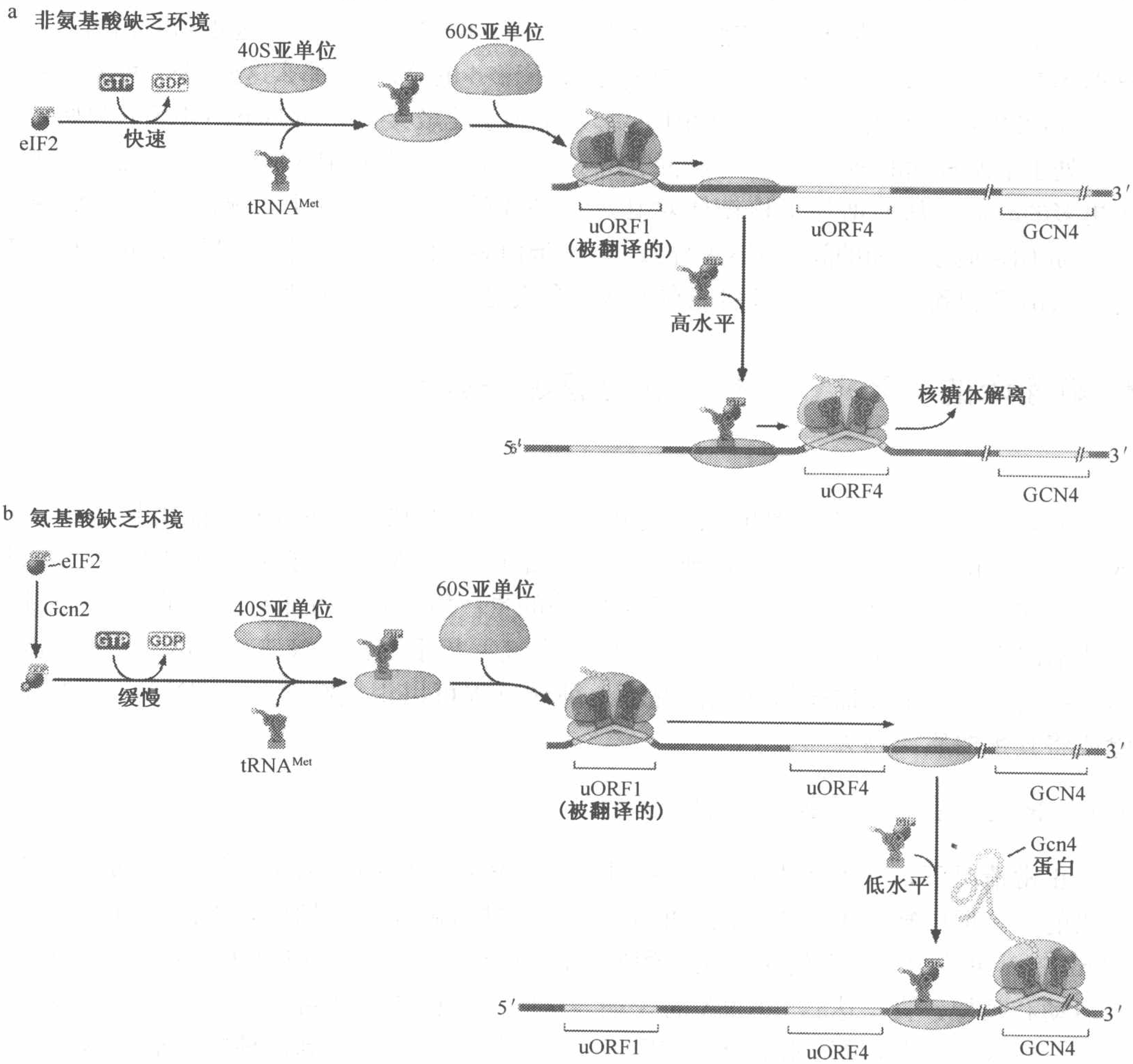


图14-48 应答于氨基酸缺乏的Gcn4翻译调控。如正文中详细介绍的，编码酵母转录激活因子Gcn4的ORF上游有4个叫做uORF的短ORF（此处只显示uORF1和uORF4）。第一个上游ORF首先翻译，由于这一ORF的特殊性质，大约50%的40S亚基在翻译终止之后仍然结合于mRNA上，继续扫描Gcn4 mRNA。a. 当氨基酸含量丰富时，eIF2B激活eIF2使其迅速将GTP转换成GDP。这使得eIF2-GTP-Met-$tRNA_i^{Met}$与40S亚基迅速结合并具备识别另外三个ORF其中之一的能力。这些uORF中任何一个的翻译都将导致翻译的完全终止。b. 在氨基酸缺乏时，由eIF2-α激酶Gcn2介导的eIF2磷酸化降低eIF2B激活GTP结合于eIF2的能力。eIF2-GTP水平的降低导致eIF2-GTP-Met-$tRNA_i^{Met}$。与40S亚基的结合变慢。起始tRNA结合速率的降低增加了进行扫描的核糖体在能够识别AUG前通过其余uORF的可能性，因此对Gcn4的翻译有利。（经得同意后，对Hinnebusch A. G. et al. 1997. *J，Biol. Cell.* 272：21661-21664的图1进行了修改 © American Society for Biochemistry & Molecular Biology.）

在氨基酸缺乏的情况下，几个事件的综合效应降低了 TC 与 40S 亚基结合的速率。有限的氨基酸导致空载 tRNA 的增多，这反过来激活一个叫做 Gcn2 的 eIF2 α 激酶。正如在 eIF2 α 激酶介导的翻译总体调控中所介绍的，当 Gcn2 磷酸化 eIF2 之后，其与 GTP 结合的能力降低。由于 eIF2 只有在 GTP 存在的情况下才与 Met-$tRNA_i^{Met}$ 结合，这导致胞内 TS 数量减少并降低 TC 结合 40S 亚基的速率（图 14-48b）。结合速率的降低意味着 40S 亚基在缺乏发现 AUG 的能力的情况下，沿 mRNA 序列进行更远距离的扫描。如果核糖体在与 eIF2-$tRNA^{Met}$重新结合之前过了 uORF2-4 的 AUG，那么这些 ORF 就不会被翻译。uORF2-4 的起始密码子与 uORF1 的起始密码子比较接近，而 Gcn4 ORF 的起始密码子却处于下游较远的地方。这一额外的距离为 TC 在小亚基碰到 ORF 之前与之结合提供了更多的时间，从而增加了核糖体对其进行翻译的机会。实际上，删除位于 uORFs 和 Gcn4 起始密码子之间的间隔 RNA 导致 Gcn4 蛋白表达的减少。因此，当有限的 TC 存在时，Gcn4 得到翻译，并能够开启能在细胞中合成更多氨基酸的基因。

依赖翻译的 mRNA 和蛋白质的稳定性调节

mRNA 会以一定的频率出现突变和损伤。这样的缺陷 mRNA 可能由转录错误造成，也可能由其合成后的损伤造成。例如，由于 mRNA 是单链的，因此更易于断裂。这样的缺陷 mRNA 可能产生不完整和不正确的蛋白质，它们可能给细胞带来负面作用。某些情况下，如发生只改变一个氨基酸的点突变，是很难消除突变的 mRNA 或其蛋白质产物的。但在下面描述的其他情况下，翻译过程可以识别缺陷的 mRNA，并予以去除或将其蛋白质产物去除。

SsrA RNA 援救翻译缺陷 mRNA 的核糖体

正常情况下，一个终止密码子对 mRNA 从核糖体中的脱离是必需的。那么，对在适当的可读框中缺失了终止密码子的 mRNA 上进行翻译的核糖体，将怎么办呢？这样的 mRNA 可能由不完全转录或核酸酶的作用而产生。这一类型的 mRNA 的翻译可以正常启动和延伸，直至达到 mRNA 的 3′端。这时，核糖体不能再移动了。此处也没有与氨基酰tRNA或释放因子相结合的密码子。如果没有一定的机制使核糖体脱离这些缺陷 mRNA，那么很多核糖体将永久地受困，不能参与多肽的合成。在原核细胞，这样的核糖体是由一种嵌合的 RNA 分子解救的，这种 RNA 分子部分是 tRNA，部分是 mRNA，称为 tmRNA。

SsrA RNA 是一个由 457 个核苷酸组成的 tmRNA，其 3′端的区域与$tRNA^{Ala}$非常相似（图 14-49）。这一相似性使得 SsrA RNA 可以负载 Ala 并结合 EF-Tu-GTP。当一个核糖体停止于mRNA的 3′端时，$SsrA^{Ala}$-EF-Tu-GTP 结合于核糖体的 A 位点并参与肽转移酶反应，就如其他的 tRNA 一样。肽酰 SsrA RNA 的易位导致缺陷的 mRNA 从核糖体中释放出来。值得注意的是，SsrA RNA 的易位也导致这一 RNA 的一部分进入核糖体的 mRNA 结合隧道。SsrA RNA 的这一部分将不完全的 mRNA 的可读框延伸了

10个密码子，并在其后紧跟着一个终止密码子。

SsrA RNA结合的最终结果是，当缺陷mRNA从核糖体中释放出来时，不完整的多肽的碳基端融合了一个10个氨基酸的"标签"，而核糖体重新进入翻译的循环过程中。有趣的是，这种10个氨基酸的"标签"被细胞的蛋白水解酶识别，蛋白水解酶很快地将这一标签以及它附着的多肽降解。这样，缺陷mRNA的多肽产物很快就被清除，从而防止了其对细胞可能带来的伤害。

SsrA RNA为什么只能与不变的核糖体结合呢？这是因为它的巨大体积（它比标准的tRNA大4倍以上），在正常的延伸过程中它不能结合于A位点。相反，当mRNA的3′端缺失时，A位点就有更大的体积来容纳较大的RNA（图14-49）。因此，只有核糖体停止于mRNA的3′端才能为SsrA RNA提供潜在的结合位点。

真核细胞降解不完整的或终止密码子提前的mRNA

真核细胞的翻译与mRNA的衰减过程紧密相连（图14-50a）。这一连锁是由两种监控被翻译的mRNA的完整性的机制来实现的。例如，当一个mRNA分子含有提前的终止密码子时（称为无义密码子，见第15章），这一mRNA很快地被降解，这一过程称为无义密码子介导的mRNA衰减（nonsense mediated mRNA decay，图14-50b）。哺乳动物细胞中，对含有提前的终止密码子的mRNA的识别依赖于聚集在mRNA的可读框内的蛋白复合体。这些外显子结合处的复合体是作为剪接的结果聚合在mRNA

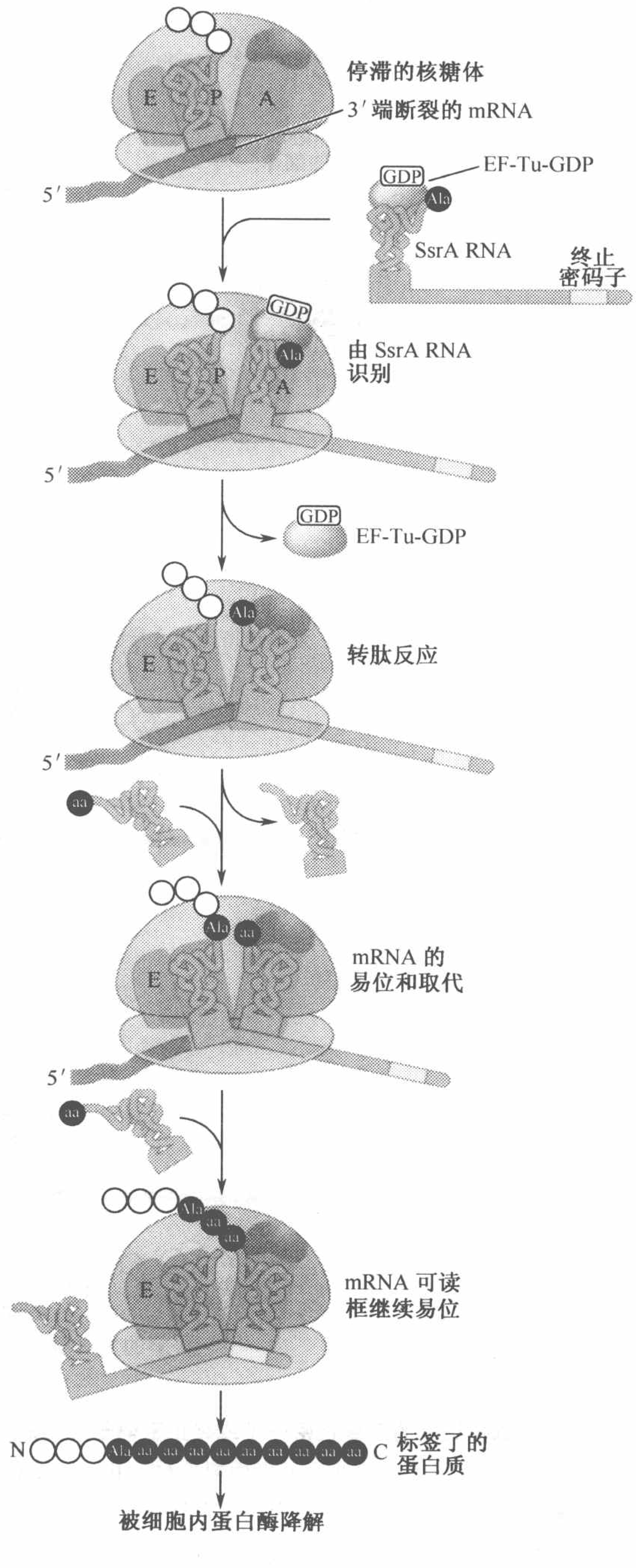


图14-49 tmRNA SsrA援救停滞于未成熟而终止的mRNA上的核糖体。 SsrA模仿tRNA，但只能结合停滞于mRNA 3′端的核糖体。一旦结合，SsrA的一部分序列作为一个新的"mRNA"工作。

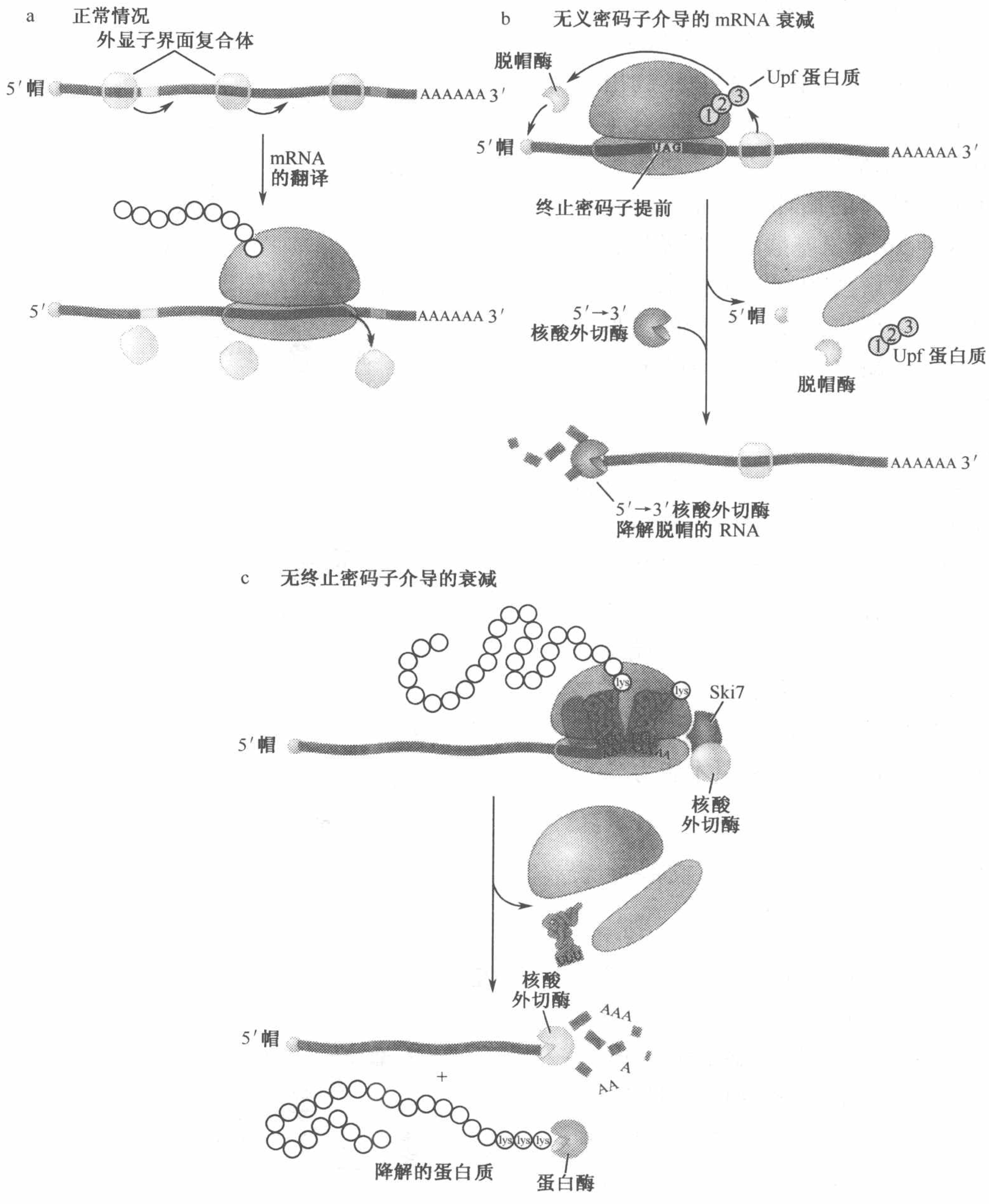


图 14-50 真核细胞的无终止密码子或终止密码子提前的 mRNA 是降解的靶标。 a. 正常 mRNA 的翻译（核糖体）移动取代了所有的外显子界面复合体。b. 无义密码子介导的衰减。终止密码子提前的 mRNA 的翻译：（核糖体）移动不能取代一个或多个外显子界面复合体，导致 Upf1、Upf2 和 Upf3 结合到核糖体，一旦与核糖体结合，这些蛋白质激活一种脱帽酶，将 mRNA 的 5′帽去除。没有 5′帽的 mRNA 很快被 5′→3′核酸外切酶降解。该酶由于 5′帽的存在无法降解 mRNA。c. 无终止密码子介导的衰减。在终止密码子缺失的情况下，多聚 A 尾得到翻译。含有 Ski7 和 3′→5′核酸外切酶的复合体结合于停滞在多聚 A 尾 3′端的核糖体。导致核糖体从 mRNA 上释放以及 mRNA 的降解。同 SsrA 介导的无终止密码子衰减一样，由多聚 A 尾翻译的位于蛋白末端的多聚 lys 是蛋白质降解的靶子。

上的，并正好坐落于每一外显子-外显子结合处的上游（第 13 章）。一般来讲，当第一个核糖体翻译某个 mRNA 时，这些复合体在 mRNA 进入核糖体的解码中心时被置换出来。但是，如果 mRNA 存在一个提前的终止密码子（源于基因的突变或转录与剪接过程中的错误），那么核糖体就在这些复合体被置换之前，从mRNA上脱离出来。在这些情况下，复合体作用于提前终止的核糖体，这样就激活一种去除 mRNA5′端帽结构的酶。由于 mRNA 通常就是受 5′端帽结构保护来防止降解的，因而帽结构的去除就导致mRNA很快地被5′→3′核酸外切酶所降解。

另一种机制拯救翻译缺乏终止密码子的 mRNA 的核糖体，称为无终止密码子介导的 mRNA 衰减（nonstop mediated decay，图 14-50c）。与原核细胞不同，真核细胞的 mRNA 以多聚 A 尾结束。当缺失终止密码子的 mRNA 翻译时，核糖体就会翻译多聚 A 尾（因为没有终止密码子在多聚 A 尾之前终止翻译）。这就导致了蛋白质的末端添加了多个赖氨酸（AAA 是赖氨酸的密码子），并使核糖体停止在 mRNA 的 3′端。停止的核糖体与 Ski7 蛋白质（与Ⅱ类释放因子 eRF3 相关）结合，刺激了核糖体的解离并募集3′→5′核酸外切酶降解“无终止”mRNA。另外，在碳基末端含有多聚赖氨酸的蛋白质是不稳定的，导致源自无终止密码子的 mRNA 的蛋白质快速降解。这样与原核细胞一样，由缺失终止密码子的 mRNA 翻译的蛋白质很快地从细胞中去除。

无义密码子介导和无终止密码子介导的 mRNA 衰减都有一个引人注目的特性：mRNA 的降解都需要受损 mRNA 的翻译。在没有翻译的情况下，受损的 mRNA 并不被很快地降解，而且具有正常的稳定性。因此，虽然不是直接的，真核细胞要依赖翻译的机制来校正它们的 mRNA。

小结

蛋白质是依据信使 RNA（mRNA）的模板在翻译过程中合成的。翻译需要将核苷酸序列的信息解码为线性多肽链的氨基酸序列。蛋白质合成的机器由 4 个主要的部分组成：mRNA；连接 RNA，也称为运输 RNA，即 tRNA；将氨基酸连接于 tRNA 的氨基酰 tRNA 合成酶；以及核糖体，它是蛋白质和 RNA 组成的多亚基复合体，催化肽键的形成。

mRNA 含有蛋白质的编码序列以及翻译起始和终止的识别单元。编码序列称为可读框（ORF），由一系列称为密码子的、相连的三核苷酸单位组成。一个 ORF 确定一条多肽链。每个 ORF 都起始于一个起始密码子并结束于一个终止密码子。起始密码子在原核细胞中通常是 AUG 或 GUG，而在真核细胞总是 AUG。原核细胞中，起始密码子的前面有一段与核糖体的 16S RNA 互补的序列，它负责将核糖体排列在起始密码子上。在真核细胞，mRNA 的 5′端含有一个称为 5′帽子的特殊结构，它负责募集核糖体。真核的 mRNA 以一串 A 结束，称为 poly-A 尾，它增强翻译的效率。原核的 mRNA 经常含有两个或多个可读框，称为多顺反子。真核的 mRNA 一般只包含一个可读框，称为单顺反子。

tRNA 是 mRNA 的密码子和准备加入多肽链的氨基酸之间的物理界面。tRNA 是 L-型的分子，其一端（5′端）为反密码子环，3′端为凸出的 5′-CCA-3′序列。反密码子与密码子互补，通过碱基配对识别。其 5′-CCA-3′端是氨基酸的结合位点，氨基酸的加入是通过氨基酸的碳基和 3′端核糖的羟基形成的酰基来实现的。

氨基酰 tRNA 合成酶分两步将氨基酸结合于 tRNA，这一过程称作负载。一个氨基酰 tRNA 合成酶负责某种氨基酸的所有 tRNA 的负载。合成酶通过与 L-型 tRNA 的两端作用来识别正确的 tRNA。合成酶负责将正确的氨基酸连接到相关的 tRNA 上，而且这一过程具有很高的保真度。一些氨基酰 tRNA 合成酶通过校正机制来获得更高的准确率。

核糖体由大、小亚基组成，大亚基含有肽键形成的位点（肽酰转移酶中心），小亚基含有解码 mRNA 的位点（解码中心）。每一亚基由一种或多种 RNA 和多种蛋白质组成。RNA 不仅是亚基的主要结构组分，而且负责核糖体的主要功能。完整的核糖体含有 3 个横跨两个亚基的 tRNA 结合位点：A 位点是负载 tRNA 进入核糖体的位置，P 位点含有肽酰 tRNA，而 E 位点是脱氨基酸的 tRNA 离开的位置。

一个蛋白质的翻译包括小、大亚基的结合与分离的循环。在核糖体循环中，小、大亚基在可读框的起始点结合起来，并在 ORF 翻译完成后分离开来。mRNA 的翻译始于 ORF 的 5′端，多肽链的合成以氨基端向碳基端的方向进行。

翻译有 3 个主要的步骤：起始、延伸和终止。原核细胞中，起始需要通过（mRNA 的）核糖体结合位点（RBS）与（小亚基的）16S RNA 的作用来将核糖体吸引到 mRNA 上。这一作用由 3 个辅助蛋白（称为起始因子 IF1、IF2 和 IF3）协助，使两个亚基保持分离，并将一个特殊的起始 tRNA 携带至起始密码子。起始 tRNA 的反密码子和起始密码子的配对引发了大亚基的结合、起始因子的释放以及起始 tRNA 进入 P 位点。这就是原核细胞的起始复合体，它为接受一个负载 tRNA 进入 A 位点并进行第一个肽键的形成做好了准备。

真核细胞的 mRNA 通过 5′帽的识别和许多辅助因子的作用来吸引小亚基。然后小亚基以 5′→3′方向沿着 mRNA 扫描，直到遇见 AUG——起始密码子。与原核细胞一样，只有在起始密码子 AUG 被识别后大亚基才能结合于 mRNA。

翻译延伸过程的第一步是将一个负载 tRNA 引入 A 位点。这在原核细胞是由 GTP 结合蛋白 EF-Tu 来催化的。在真核细胞则由它的等价物来催化。在氨酰基团进入肽酰转移酶中心之前，有许多机制保证了密码子和反密码子之间的正确配对。然后，通过将肽链从 P 位点的 tRNA 转移到 A 位点的氨基酰 tRNA 来形成肽键。肽键的形成是由大亚基的肽转移酶中心的 RNA 及 P 位点 tRNA 的 3′OH 催化的。这一核糖酶促使 A 位点的氨酰 tRNA 的氨基基团亲核进攻 P 位点上连接延伸着的多肽链的 tRNA 的碳基基团。最后，核糖体易位至下一个空的密码子上，这一过程由肽转移酶和延伸因子 EF-G（或真核的等价物）共同催化。易位的结果是 P 位点的脱氨基酸的 tRNA 移入 E 位点，并从 E 位点离开核糖体；A 位点的肽酰 tRNA 进入刚刚空出的 P 位点。mRNA 的下一个密码子移入空的 A 位点，并做好接受由 EF-Tu 携带的另一个负载 tRNA 的准备。

翻译终止于核糖体遇见一个终止密码子，（对终止密码子的）识别是通过原核的两个 Ⅰ 类释放因子之一和真核的一个 Ⅰ 类释放因子进行的。释放因子刺激肽酰 tRNA 的

水解，导致完整的肽链释放出来。最后，一个Ⅱ类释放因子、一个核糖体循环因子和一个起始因子（原核生物中为 IF3）通过引起 mRNA 和脱氨基酸的 tRNA 的释放以及核糖体解离为大、小亚基完成翻译的终止过程。核糖体循环终于完成，小亚基又准备开始新一轮的多肽合成。

很多基因的表在是在转录水平受调控的；但是一些基因也在翻译起始水平受到调控。在细菌细胞中，这种调控总体上通过抑制小亚基与 RBS 的结合来进行。这种抑制可由蛋白或 RNA 结合到 RBS 附近的 mRNA 序列上来介导。真核生物翻译的总体水平由 4E-BP 蛋白及 eIF2α 激酶来调控，前者可结合 eIF4E 并且与 eIF4G 竞争与之结合的能力（注：原文有误），后者抑制 eIF2 与 GTP 结合的能力。特殊的真核 mRNA 的翻译调控有时由小 uORFs 介导，它限制核糖体小亚基接近下游的 ORF。在 Gcn4 的翻译中，细胞利用 eIF2 · Met-$tRNA_i^{Met}$ 对氨基酸水平的依赖调控 Gcn4 的翻译。只有当 eIF2 · Met-$tRNA_i^{Met}$ 有限时，才能使处于扫描状态的小亚基在没有解离前遇到 Gcn4 起始密码子。

细菌和真核细胞都采用翻译来监控 mRNA 的完整性并去除不完整的 RNA 及其蛋白产物。缺失终止密码子的 mRNA 将导致合成被胞内蛋白酶识别并降解的异常蛋白。在真核细胞中，相应的 mRNA 也被降解。与之相似，带有一个未成熟的终止密码子的真核 mRNA 因为核糖体无法置换结合于 mRNA 的蛋白质而被识别。如果这些蛋白质不被置换，它们就募集指导 mRNA 降解的酶。

（吴东颖　郭需石　译　吴琳　苏夜阳　校）

参考文献

书

Mathews M.B., Sonenberg N., and Hershey J.W.B., eds. 2007. *Translational control in biology and medicine.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

tRNA及氨酰tRNA合成酶

Arnez J.G. and Moras D. 1997. Structural and functional considerations of the aminoacylation reaction. *Trends Biochem. Sci.* **2:** 189–232.

Ibba M. and Soll D. 2000. Aminoacyl-tRNA synthesis. *Annu. Rev. Biochem.* **69:** 617–650.

核糖体

Koronstelev A., Trakhanov S., Laurberg M., and Noller H.F. 2006. Crystal structure of a 70*S* ribosome-tRNA complex reveals functional interactions and rearrangements. *Cell* **126:** 1065–1077.

Moore P.B. and Steitz T.A. 2005. The ribosome revealed. *Trends Biochem. Sci.* **30:** 281–283.

Poehlsgaard J. and Douthwaite S. 2005. The bacterial ribosome as a target for antibiotics. *Nat. Rev. Microbiol.* **3:** 870–881.

Ramakrishnan V. 2002. Ribosome structure and the mechanism of translation. *Cell* **108:** 557–572.

Rodnina M.V., Beringer M., and Wintermeyer W. 2006. How ribosomes make peptide bonds. *Trends Biochem. Sci.* **32:** 20–26.

Selmer M., Dunham C.M., Murphy F.V., Weixlbaumer A., Petry S., Kelley A.C., Weir J.R., and Ramakrishnan V. 2006. Structure of the 70*S* ribosome complexed with mRNA and tRNA. *Science* **313:** 1935–1942.

翻译

Broderson D.E. and Ramakrishnan V. 2003. Shapes can be seductive. *Nat. Struct. Biol.* **10:** 78–80.

Laursen B.S., Sorenson H.P., Mortenson K.K., and Sperling-Peterson H.U. 2005. Initiation of protein synthesis in bacteria. *Microbiol. Mol. Biol. Rev.* **60:** 101–123.

Nilsson J. and Nissen P. 2005. Elongation factors on the ribosome. *Curr. Opin. Struct. Biol.* **15:** 349–354.

Nissen P., Kjeldgaard M., and Nyborg J. 2000. Macromolecular mimicry. *EMBO J.* 19: 489–495.

Weinger J.S., Parnell K.M., Forner S., Green R., and Strobel S.A. 2004. Substrate-assisted catalysis of peptide bond formation by the ribosome. *Nat. Struct. Mol. Biol.* **11:** 1101–1106.

翻译的调控

Gebauer F and Hentze M.W. 2004. Molecular mechanisms of translational control. *Nat. Rev. Mol. Cell Biol.* **5:** 827–835.

Richter J.D. and Sonenberg N. 2006. Regulation of cap-dependent translation by eIF4E inhibitory proteins. *Nature* **433:** 477–480.

第15章 遗传密码

本章概要

•

遗传密码具简并性

•

遗传密码遵循的三条规律

•

抑制突变可以存在于相同或不同的基因中

•

遗传密码几乎通用

中心法则最核心的内容是遗传信息从由4个字母组成的线性多核苷酸链传递给由20个氨基酸语言组成的多肽链。如前所述，将遗传信息传递给氨基酸序列的翻译过程发生在核糖体上，并且是由一种称为tRNA的特殊转配器分子所介导。这些tRNA能够识别由3个连续的核苷酸组成的密码子。由于每个位置上的核苷酸只有4种可能，所以三联密码子排列的总数是64（4×4×4），远远超过了氨基酸的数目。那么每一种氨基酸是由哪个三联密码子决定的呢？使用的规则又如何呢？在本章我们将要讨论遗传密码的本质、内在联系、遗传密码如何被“破译”，以及突变对mRNA编码能力的影响。

遗传密码具简并性

表15-1罗列了64种密码子排列。三联密码子5′端碱基列在表的左侧，中间的碱基排在顶上，而3′端的碱基则位于右侧。遗传密码最突出的特征之一是64个密码子中的61个分别定义一个氨基酸，而其余的3种则是氨基酸链终止信号（见下面）。这意味着很多氨基酸是由超过1个以上的密码子定义的，这种现象称为密码子的简并性（degeneracy)。定义同一个氨基酸的几个密码子称为同义密码子（synonym)。如UUU和UUC是Phe的同义密码子，而Ser是由同义密码子UCU、UCC、UCA、UGG、AGU和AGC编码的。事实上，当密码子的第1和第2位核苷酸完全相同时，第3位核苷酸无论是胞嘧啶还是尿嘧啶，均编码相同的氨基酸，通常腺嘌呤和鸟嘌呤可以相似地进行互换。然而，并不是所有简并密码子的前两位核苷酸都是相同的。例如，Leu既可以由UUA和UUG编码，也可以由CUU、CUC、CUA和CUG编码（图15-1）。很多生物基因组DNA分子中AT/CG的比例有很大的变异，但它们的蛋白质分子中氨基酸的比例却没有相应的大变化，这一现象可以用密码子的简并性，特别是在密码子第3位单核

苷酸的常见的胞嘧啶与尿嘧啶或者腺嘌呤与鸟嘌呤的互换性来解释。(例如，某些细菌的基因组中 AT/GC 的比例差别极大，但又紧密相关足以编码有着高度相似的氨基酸序列的蛋白质。)

表 15-1 遗传密码

第 1 个位置(5′端)	第 2 个位置 U		C		A		G		第 3 个位置(3′端)
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
	UUC		UCC		UAC		UGC		C
	UUA	Leu	UCA		UAA*	终止子	UGA*	终止子	A
	UUG		UCG		UAG*	终止子	UGG	Trp	G
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	Gln	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	Lys	AGA	Arg	A
	AUG†	Met	ACG		AAG		AGG		G
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	Glu	GGA		A
	GUG		GCG		GAG		GGG		G

* 肽链“终止密码子”或者“‘无义’密码子”。

† 细菌还用来定义起始的甲基-Met-$tRNA^{fMet}$。

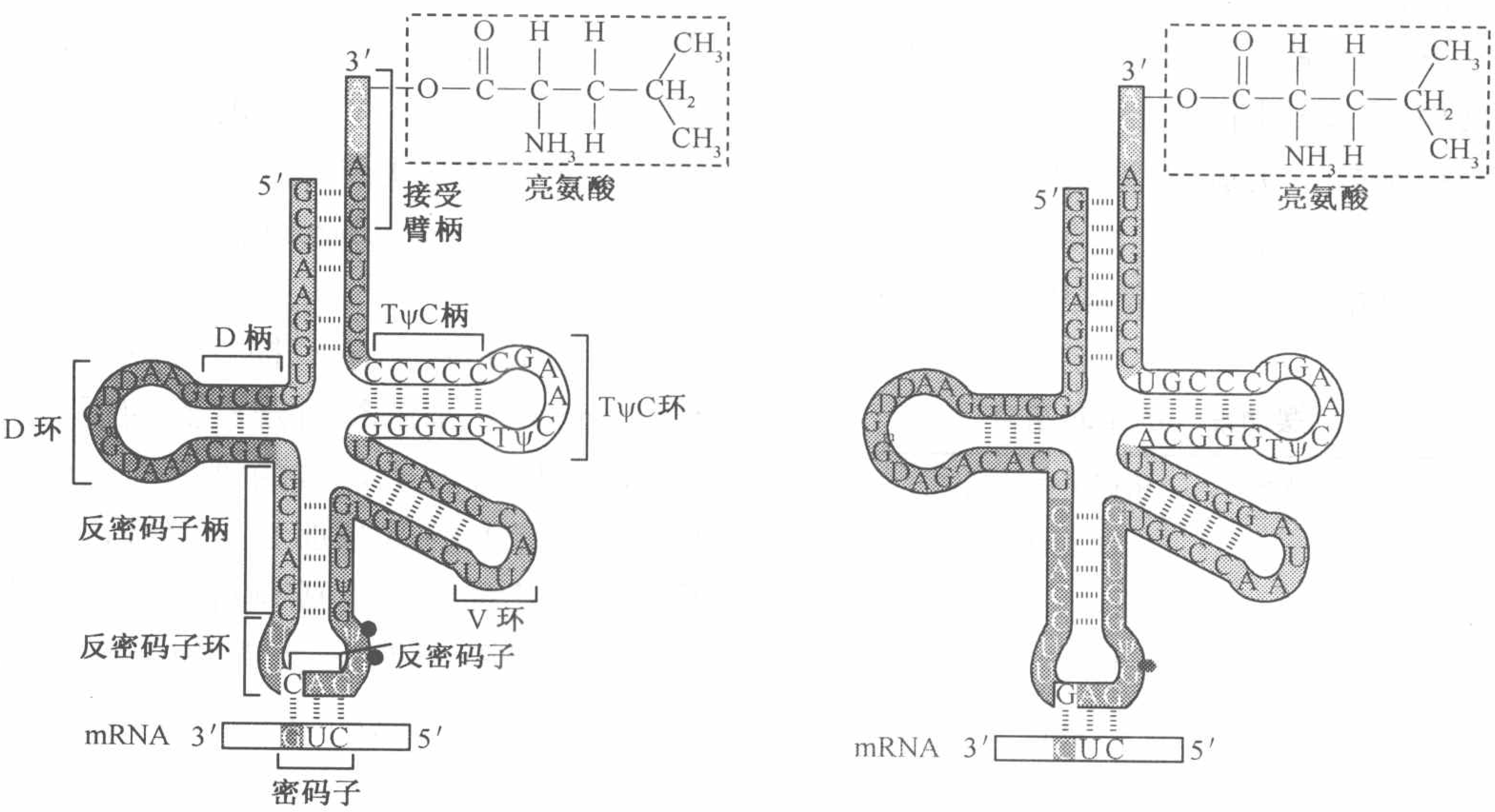


图 15-1 两个 $tRNA^{Leu}$分子的密码子-反密码子配对。标出了 tRNA 结构中关键的柄、环区域(第 14 章)。与 G(反密码子的 3′端)连接的红色六边形表示碱基 N1 位置的甲基化。注意密码子的方向为 3′→5′。

认识密码子组成的规律

考察密码子在遗传密码中的分布，可以得到遗传编码的进化趋向于将突变的有害效应减少到最小的规律。例如，密码子第 1 个位置上核苷酸的突变将使编码的氨基酸即使不完全相同，也比较相似。另外，第 2 个位置的核苷酸为嘧啶的密码子所编码的氨基酸往往呈疏水性，而第 2 个位置的核苷酸为嘌呤的密码子所编码的蛋白质往往是极性氨基酸（表 15-1 和第 5 章图 5-4）。由于转换（A∶T→G∶C 或者 G∶C→A∶T 替换）是最常见的点突变类型，密码子第二个位置的核苷酸变化往往使一个氨基酸被另一个非常相似的氨基酸所取代。最后，如果密码子第 3 个位置上的核苷酸发生转换突变，也极少产生出不同的氨基酸。在半数情况下，即使在这个位置上发生颠换突变也不会带来什么后果。

遗传密码中另一个显而易见的规律是：当密码子的第 1 个和第 2 个位置均为 G 或 C 时，第 3 个位置上的核苷酸无论为 A、T、C 和 G 中的哪一个，均编码相同的氨基酸（如 Pro、Ala、Arg 或者 Gly）。另一方面，当密码子的第 1 个和第 2 个位置均为 A 或 U 时，第 3 个位置的核苷酸对蛋白质的编码就会产生影响。因为 G∶U 碱基对比 A∶U 碱基对强，当密码子的第 1 个和第 2 个位置为结合力强的 G∶C 碱基对时，密码子将能够耐受第 3 个位置上碱基的错配。因此，在第 3 个位置上使所有 4 种核苷酸定义相同的氨基酸可能已逐渐进化为一种安全机制，以便在阅读这种密码子时将错误减少到最小。

反密码子的摆动配对

最初人们认为每一个密码子都有特异的 tRNA 反密码子。如果这种假说是正确的，生物体中应该有至少 61 种不同的 tRNA，可能还有另外 3 种对应于链终止密码子。然而有证据表明：某种高度纯化的已知序列的 tRNA 可以识别几种不同的密码子。也有研究发现反密码子除了 4 种常规的碱基之外，还有第 5 种碱基次黄嘌呤（inosine，I）。和其他 tRNA 中的次要碱基一样，次黄嘌呤是通过对完整 tRNA 链中的碱基进行酶修饰而得来的。次黄嘌呤源于腺嘌呤，即腺嘌呤分子上第 6 个碳原子脱去氨基，成为次黄嘌呤的 6-酮基基团。[实际上，次黄嘌呤是由核糖和次黄嘌呤（hypoxanthine）碱基组成的核苷酸，但现已常常用来指碱基，就如同我们这里所用的。]

表 15-2 摆动理论的配对组合

反密码子中的碱基	密码子中的碱基
G	U 或 C
C	G
A	U
U	A 或 G
I	A、U 或 C

1966 年 Francis Crick 提出了摆动理论（wobble concept）来解释所观察到的现象。摆动理论指出，反密码子 5′端的碱基在空间上不如其他 2 个位置上的碱基那么受限制，可以和密码子 3′端的几种碱基形成氢键。不是所有组合都行，配对局限于表 15-2 中的碱基。例如，位于摆动位置的 U 可以和 A 或 G 配对，I 可以和 U、C 或 A 配对(图15-2)。摆动规则所允许的配对是那些核糖与核糖之间的距离与标准的 A∶U或G∶C碱基对相似的配对。(除了 I 和 A 碱基对）嘌呤与嘌呤、嘧啶与嘧啶的配对分别使核糖与核糖之间的距离过长或者过短。

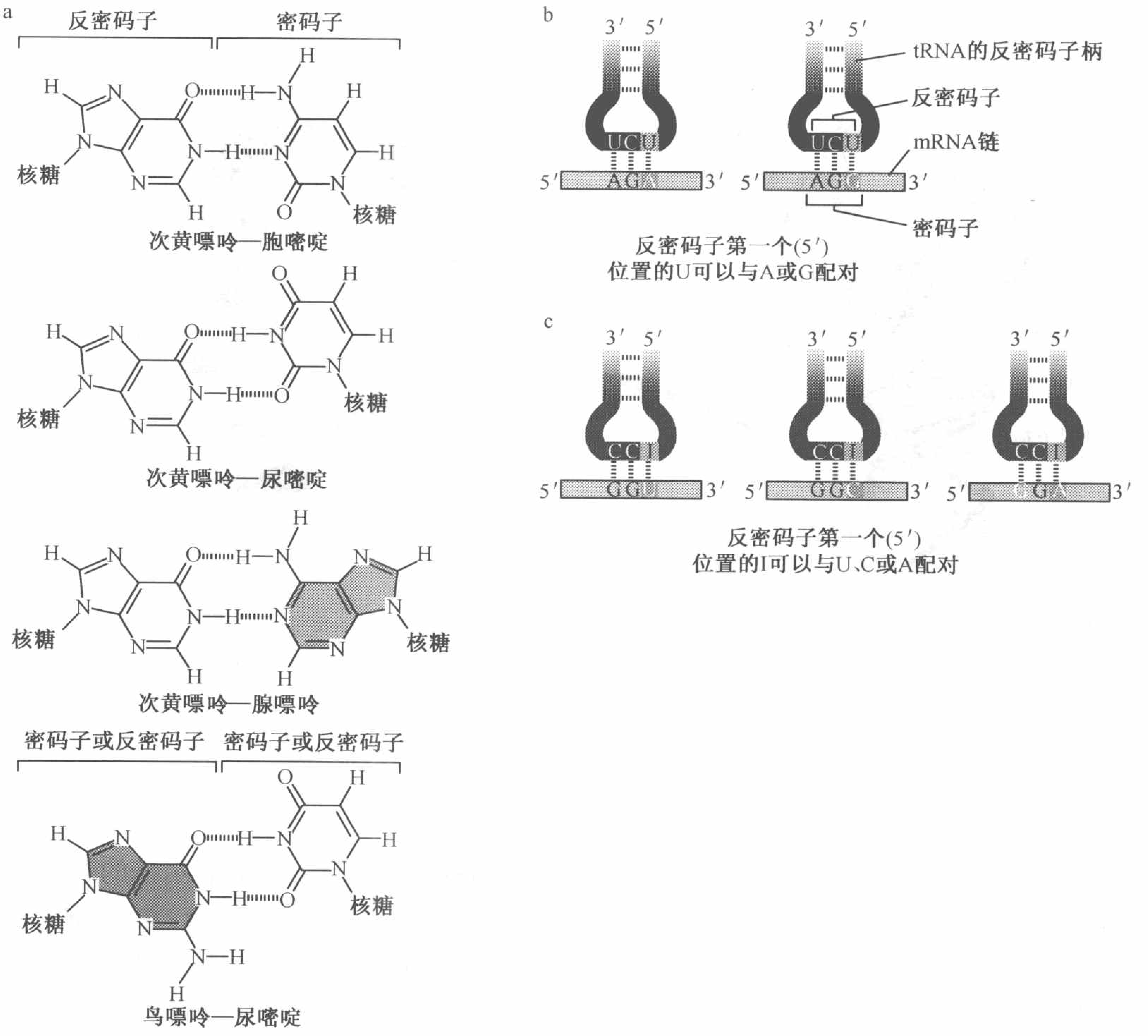


图 15-2　摆动碱基配对。注意所有摆动碱基的核糖与核糖之间的距离与 A：U 或 G：C 标准碱基对的距离相近。

摆动规则不允许单个的 tRNA 分子识别 4 个不同的密码子。3 个密码子的识别只有当反密码子的第一个（5′）位置是 I 时才行。

自 1966 年以来所积累的几乎所有的实验证据均支持摆动理论。例如，摆动理论正确地预测到至少有 3 种 tRNA 存在，以识别 Ser 的 6 种密码子（UCU、UCC、UCA、UCG、AGU 和 AGC）。其他两个由 6 种密码子编码的氨基酸（Leu 和 Arg）也存在不同的 tRNA 以识别在第 1 或者第 2 个位置上有不同核苷酸的那些密码子。

在 tRNA 的三维结构中，3 个反密码子碱基（及其反密码子环中随之相连的 2 个碱基）基本上都指向同一个方向，它们精确的构象主要由碱基平面之间的堆叠作用所决定（图 15-3）。反密码子的第一个（5′）碱基位于碱基堆叠的末端，其运动也许比其他两个碱基相对自由些，因而，在密码子的第三个（3′）碱基位置上摆动。相反，不仅反密码子的第三个（3′）碱基位于碱基堆叠的中间，而且相邻的碱基总是庞大的修饰嘌呤残基。因此，碱基运动的受限性可能解释了摆动为什么不出现在密码子的第一个（5′）位置上。

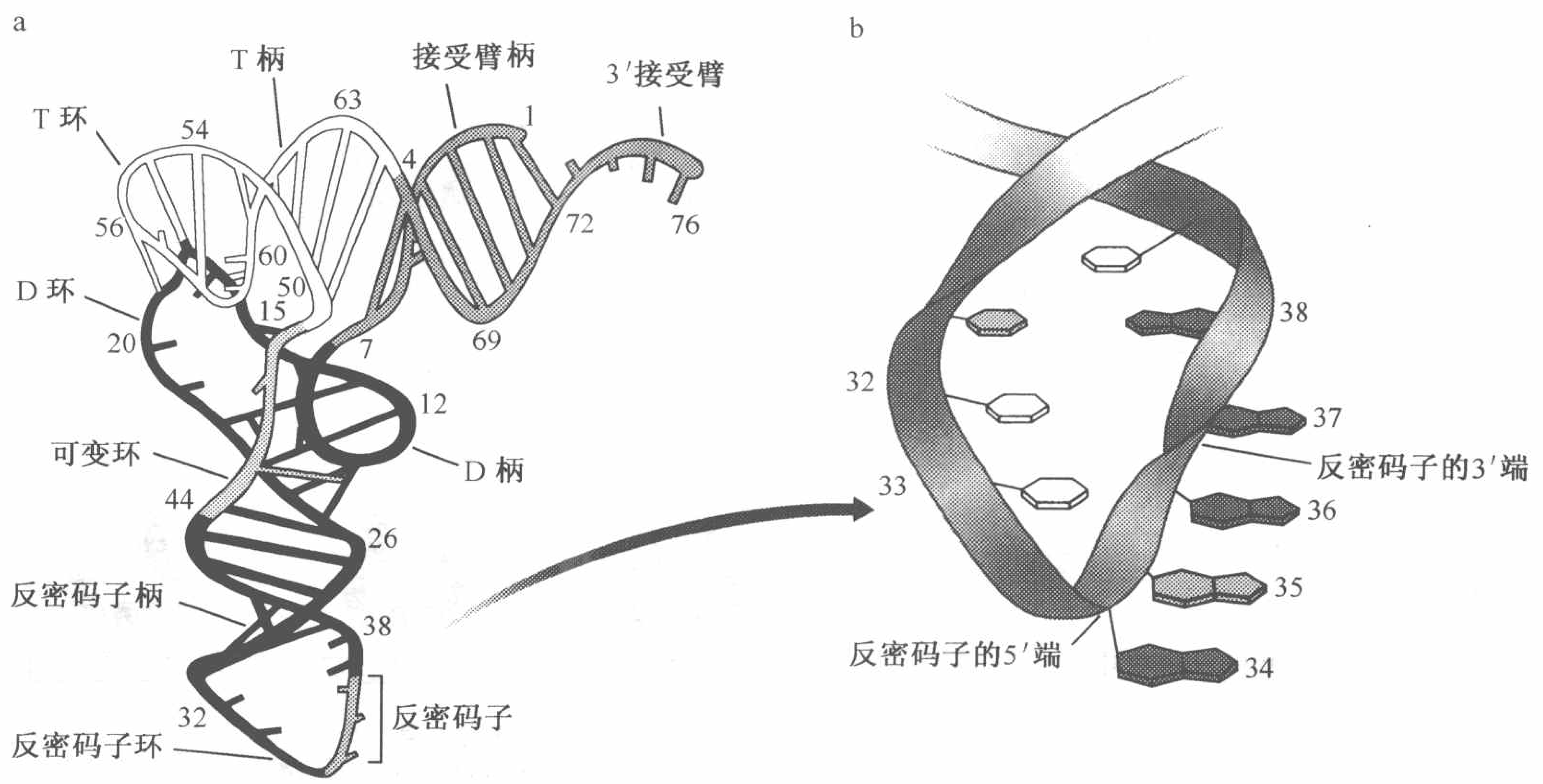


图 15-3 酵母 tRNA^{Phe}的结构。 a. 基于 X 射线衍射数据得到的 L-型分子。b. 反义密码子环的放大图。反义密码子碱基（34～36）显示为红色。反义密码子及在 3′端的随后两个碱基（37 和 38）是部分堆叠的。可以看见反义密码子 5′端的碱基比 3′端完全堆叠的碱基更自由地摆动。（来源：Kim S-H，et al. 1974，Proc. Natl. Acad. Sci. 71：4970.）

三个密码子指导链的终止

如前所述，有三个密码子并不与任何氨基酸相对应。相反，它们表示链的终止。在第 14 章我们讨论到，这些链终止密码子 UAA、UAG 和 UGA 不是由特殊的 tRNA 来阅读，而是由称为释放因子（RF）的特殊蛋白质所阅读（细菌中为 RF1 和 RF2，在真核生物中为 eRF1）。释放因子进入核糖体的 A 位点，激发位于 P 位点的肽酰 tRNA 的水解，使新合成的蛋白质释放出来。

遗传密码是如何破译的?

建立氨基酸与三联密码子之间的对应关系是分子生物学史上最伟大的成就之一（历史回顾见第 2 章）。这些对应关系是如何建立的呢？到 1960 年，人们对 mRNA 如何参与蛋白质合成的总体轮廓已经有了基本了解。然而对是否能够很快详细认识遗传密码本身并不乐观。人们相信要认识一个氨基酸的密码子，不但需要准确地知道基因的核苷酸序列，而且需要了解相应的蛋白质产物的氨基酸序列。当时，测定氨基酸序列尽管是一个十分繁琐的工作，但在实验室已经十分行之有效。而另一方面，当时通用的 DNA 测序方法还非常原始。幸运的是，这个明显的障碍并未阻止研究进程。1961 年，即在 mRNA发现后仅一年，采用人工合成 mRNA 技术和无细胞的蛋白质合成系统使遗传密码的破译成为可能（第 2 章）。

通过人工合成的 mRNA 激发氨基酸的掺入

生物化学家发现，从活跃地进行蛋白质合成的大肠杆菌细胞中得到的提取物能够促

使放射性标记的氨基酸合成蛋白质。在这些提取物中，蛋白质合成在数分钟内进行得很快，然后逐渐停止。在这段时间中，由于提取物中存在的降解酶的作用，导致 mRNA 相应减少。然而，在已经停止蛋白质合成的提取物中加入新的 mRNA，能够使蛋白质合成立即恢复。

细胞提取物依赖外在加入的 mRNA 增加蛋白质的合成，为利用合成的多聚核糖核酸来解开遗传密码的性质提供了机会。这些合成的模板是利用多核苷酸磷酸化酶而产生的，该酶催化如下的反应：

$$[XMP]_n + XDP \rightleftharpoons [XMP]_{n+1} + P \qquad \text{(反应式 15-1)}$$

式中：X 表示碱基，$[XMP]_n$ 表示有 n 个核苷酸的 RNA。

多核苷酸磷酸化酶的正常作用是分解 RNA，在生理条件下偏向于将 RNA 降解成二磷酸核苷。然而，在高浓度的二磷酸核苷存在下，多核苷酸磷酸化酶能够从 3′→5′方向在核苷酸之间催化磷酸二酯键的形成，从而合成 RNA 分子（图 15-4）。多核苷酸磷酸化酶催化合成 RNA 并不需要 DNA 或者 RNA 模板，所合成的产物完全依赖于反应混合物中各种二磷酸核苷的比例。例如，当只有二磷酸腺苷存在时，合成的 RNA 只有腺苷酸，因而称为多聚腺苷酸（polyadenylic acid）或 poly-A。同样也可以合成 poly-U、poly-C、poly-G。加入两种或者更多种不同的二磷酸核苷，可以形成复合的共聚物，如 poly-AU、poly-AC、poly-CU 和 poly-AGCU。在所有这些混合共聚物中，碱基的序列基本上是随机的，最相邻的碱基频率完全是由反应物的相对浓度决定的。例如，在 A 是 U 的 2 倍的情况下，poly-AU 分子中有类似 UAAUAUAAAUAAUAAAAUAUU……的序列

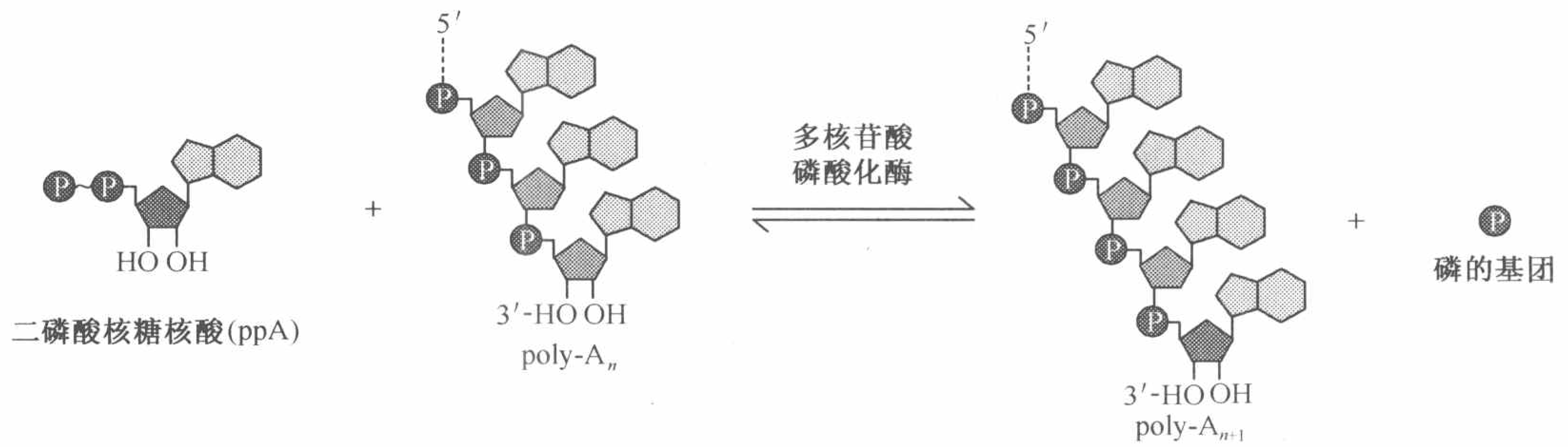


图 15-4 多核苷酸磷酸化酶反应。本图显示由多核苷酸磷酸化酶催化的多聚腺苷酸合成或降解的可逆反应。

poly-U 编码多聚 Phe

在合适的离体条件下，几乎所有合成的多聚物会与核糖体连接并作为模板。幸运的是，早期的实验使用了高浓度的 Mg^{2+}。高浓度的 Mg^{2+} 绕开了对起始因子和特殊的起始 fMet-tRNA 的需求，使 mRNA 在没有适当信号的情况下就能够开始多核苷酸链的合成。poly-U 是最先发现的具有 mRNA 活性的合成多聚核苷酸。它特异性地选择苯丙氨酰 tRNA 分子，因而形成的多肽链只含有 Phe（多聚 Phe）。因此我们知道编码 Phe 的密码子是由 3 个鸟苷酸——UUU 组成的（如第 2 章和第 21 章及下文所讨论的，一个密

码子由 3 个单核苷酸组成的观点来源于遗传学实验)。通过使用 poly-A 和 poly-C 的类似实验，发现了 CCC 是 Pro 的密码子，AAA 是 Lys 的密码子。不幸的是，此类实验无法确定 GGG 定义何种氨基酸。poly-G 中的胞嘧啶残基相互之间形成牢固的氢键，形成多链的三螺旋结构，不能和核糖体结合。

混合多聚核苷酸使其他密码子的组成得以确定

poly-AC 分子可以含有 8 种不同的密码子：CCC、CCA、CAC、ACC、CAA、ACA、AAC 和 AAA，其比例随着共聚物的 A/C 比例不同而有差异。当 AC 共聚物连接到核糖体上，引起掺入的氨基酸有 Asp、Gly、His 和 Thr，加上先前确定的由 CCC 编码的 Pro 和由 AAA 编码的 Lys，这些氨基酸在多肽链中的比例取决于 A/C 的比例。例如，由于 A 含量大大高于 C 的 AC 共聚体促进 Asp 的掺入远远多于 His，所以我们得出结论，Asp 由两个 A 和一个 C 编码，而 His 由两个 C 和一个 A 编码（表 15-3)。采用其他共聚体的类似实验使另外几个密码子的组成得以确定。然而，此类实验并不能显示三联密码子中核苷酸的顺序。采用随机的共聚物无法了解编码组氨酸的两个 C 和一个 A 的排列顺序是 CCA、CAC 还是 ACC。

表 15-3 掺入蛋白质的氨基酸*

氨基酸	掺入氨基酸的观察值	推断的密码子组成	三联体的计算频率				三联体的计算频率总和
			3A	2A1C	1A2C**	3C	
poly-AC（5∶1）							
Asp	24	2A1C		20			20
Glu	24	2A1C		20			20
His	6	1A2C			4.0		4
Lys	100	3A	100				100
Pro	7	1A2C，3C			4.0	0.8	4.8
Thr	26	2A1C，1A2C		20	4.0		24
poly-AC（1∶5）							
Asp	5	2A1C		3.3			3.3
Glu	5	2A1C		3.3			3.3
His	23	1A2C			16.7		16.7
Lys	1	3A	0.7				0.7
Pro	100	1A2C，3C			16.7	83.3	100
Thr	21	2A1C，1A2C		3.3	16.7		20

* 在无细胞提取物中随机加入 A 和 C 的共聚体后观察蛋白质中所掺入的氨基酸。掺入量是指单个氨基酸最大掺入量的百分比，然后用共聚物计算一个给定的密码子在多核苷酸产物中出现的频率。密码子的相对频率是指一个特定的核苷酸出现在密码子的指定位置的概率。例如，当 A/C 比是 5∶1 时，AAA/AAC 的比=5×5×5∶5×5×1=125∶25。如果我们假定 3A 密码子的频率为 100，则 2A 和 1C 密码子的频率为 20。将氨基酸掺入的频率和给定的计算的密码子出现频率作相关分析，就能得到推定的密码子的组成。（**，原文为 1A1C）

tRNA 和特定的三联密码子结合

1964 年，人们研发出一种直接确定某些三联密码子中 3 个核苷酸顺序的方法。这种方法基于如下事实：即使缺乏蛋白质合成需要的所有因子，特定的氨基酰 tRNA 分子仍然能够结合核糖体-mRNA 复合物。例如，当 poly-U 和核糖体混合时，只有苯丙氨酰 tRNA 会结合。相应地，poly-C 促进脯氨酰 tRNA 的结合。更为重要的是，这种特异的结合并不需要长片段 mRNA 分子的存在。事实上，三核苷酸和核糖体结合就足够了。三核苷酸 UUU 的加入导致苯丙氨酰 tRNA 的结合，如果加入 AAA，赖氨酰 tRNA 就特异地结合核糖体。这种三核苷酸效应的发现，为识别许多密码子内核苷酸的顺序提供了一种相对简便的方法。例如，三核苷酸 5′-GUU-3′促进缬氨酰 tRNA 的结合，5′-UGU-3′促进半胱氨酰 tRNA 的结合，5′-UUG-3′促进亮氨酰 tRNA 的结合（表 15-4）。尽管合成了所有 64 种可能的三联核苷酸，以确定每一个密码子的顺序，然而并非所有的密码子都是采用这种方法来确定的。一些三联核苷酸不如 UUU 或者 GUU 那样有效地结合核糖体，因而无法知道它们所编码的特异氨基酸。

表 15-4　氨基酰 tRNA 分子与三联核苷酸-核糖体复合物的结合

三联核苷酸						结合的氨基酰-tRNA
5′-UUU-3′	UUC					Phe
UUA	UUG	CUU	CUC	CUA	CUG	Leu
AAU	AUC	AUA				Ile
AUG						Met
GUU	GUC	GUA	GUG	UCU*		Val
UCU	UCC	UCA	UCG			Ser
CCU	CCC	CCA	CCG			Pro
AAA	AAG					Lys
UGU	UGC					Cys
GAA	GAG					Glu

* 注意这是采用此方法错误排列的密码子。

重复共聚物确定密码子

在三联核苷酸结合技术产生的同时，有机化学和酶技术正用于已知重复序列的多聚核糖核酸的合成（图 15-5）。核糖体可以在这些常规共聚物链上的任意点开始合成蛋白质；但是它们将特异的氨基酸掺入到肽链中。例如，重复序列 CUCUCUCU…是一种常见多肽的信使，在此多肽中，Leu 和 Ser 交替出现。同样地，UGUGUG…促进含有 Cys 和 Val 多肽链的合成。ACACAC…指导含有 Thr 和 His 交替的多肽链的合成。由三联核苷酸序列 AAG（AAGAAGAAG）重复所组成的共聚物指导合成 3 种多肽：poly-Lys、poly-Arg 和 poly-Glu。poly-AUC 以同样的方式，作为 poly-Ile、poly-Ser 和 poly-His 的模板（表 15-5）。其他密码子的确定是从四联核苷酸重复序列获得的。

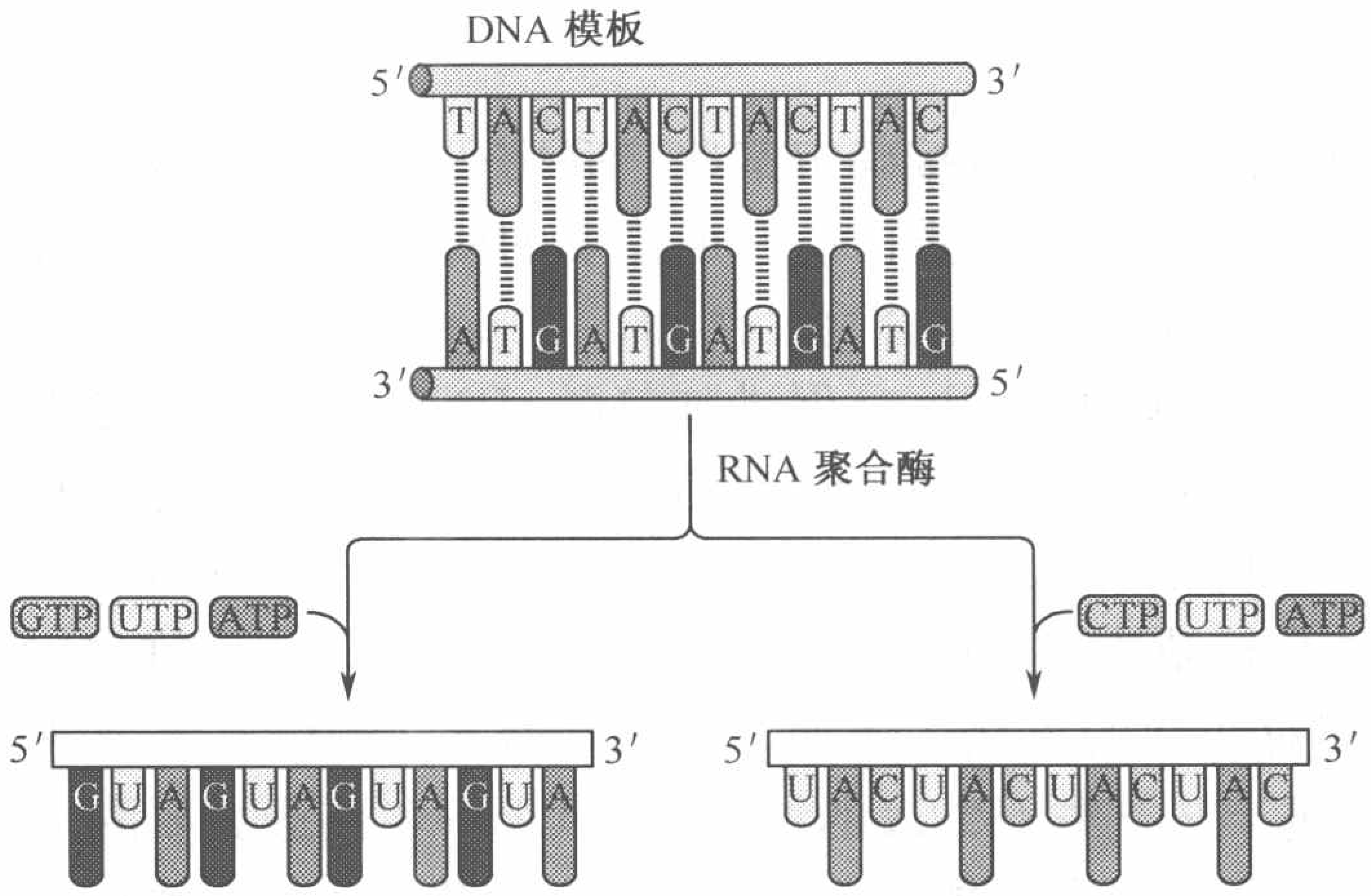


图 15-5 制备寡核苷酸。结合有机合成和 DNA 聚合酶 I 复制两种技术，可以生成含有简单重复序列的 DNA 双链。根据加入反应混合物中的三磷酸核苷的不同，RNA 聚合酶以一条或另一条 DNA 链为模板，合成不同的长链多聚核苷酸。

表 15-5 使用由 2 个或者 3 个单核苷酸合成的重复共聚物进行密码子的确定

共聚物	识别的密码子	掺入的氨基酸或者合成的多肽	确定的密码子
$(CU)_n$	CUC/UCU/CUC···	Leu	5′-CUC-3′
		Ser	UCU
$(UG)_n$	UGU/GUG/UGU···	Cys	UGU
		Val	GUG
$(AC)_n$	ACA/CAC/ACA···	Thr	ACA
		His	CAC
$(AG)_n$	AGA/GAG/AGA···	Arg	AGA
		Glu	GAG
$(AUC)_n$	AUC/AUC/AUC···	poly-Ile	AUC
	UCA/UCA/UCA···	poly-Ser	UCA
	CAU/CAU/CAU···	poly-His	CAU

所有这些实验结果加起来使得 64 种密码子中的 61 种所编码的氨基酸得到确定（表 15-1），其余 3 种是链终止密码子：UAG、UAA 和 UGA，它们不定义任何氨基酸。（注意：如前面章节所讨论，位于大肠杆菌的翻译起始序列时，AUG 作为起始密码子来指明 *N*-甲酰蛋氨酸，而不是通常的蛋氨酸。）

遗传密码遵循的三条规律

遗传密码遵循三条指导密码子在 mRNA 中的排列和使用的规律。第一条规律是密码子从 5′→3′方向阅读。例如，二聚肽 NH_2-Thr-Arg-COOH 的编码顺序可以写成：5′-ACGCGA-3′（其中 5′-ACG-3′编码 Thr，5′-CGA-3′编码 Arg），或者写成 3′-GCAAGC-

5′，即密码子按同样的顺序，但与原来的方向相反。然而，由于 mRNA 是从 5′→3′的方向进行翻译的，因此只有前者才是正确的编码序列；如果后者从 5′→3′方向进行翻译，则合成的肽链会是 NH_2-Arg-Thr-COOH，而不是 NH_2-Thr-Arg-COOH。

第二条规律是密码子不重叠，信息之间无缝隙。这意味着由相邻的三联体核苷酸所组成的密码子是连续的密码子。这样，三肽 NH_2-Thr-Arg-Ser-COOH 的编码序列是由序列 5′-ACGCGAUCG-3′中 3 个连续的、相互不重叠的三联密码子组成的。

最后一条规律是信息在固定的可读框上翻译，而可读框是由起始密码子设定的。第 14 章提到，翻译始于蛋白质编码序列 5′端的起始密码子。由于密码子是非重叠的并由 3 个连续的核苷酸组成，原则上一段核苷酸序列可以按照任何 3 个可读框进行翻译。起始密码子决定使用 3 个可能的可读框中的哪一个。例如，序列 5′… ACGACACGACGACGACG…3′可以翻译成一连串的 Thr（5′-ACG-3′）、一连串的 Arg（5′-CGA-3′）或者是一连串的 Asp（5′-GAC-3′），这些都取决于上游起始密码子的框架。

改变遗传密码的三种点突变

在研究了遗传密码的本质之后，重新回顾点突变如何引起基因编码序列改变的议题是有帮助的（参见第 9 章）。引起编码一种特异氨基酸的密码子成为编码另一种氨基酸的密码子的改变称为错义突变（missense mutation）。其结果是有错义突变的基因其蛋白质中的一个氨基酸被另一种氨基酸取代，就如经典的例子人类遗传性疾病镰状红细胞贫血，其中血红蛋白的 β-珠蛋白亚基的第 6 位 Glu 被 Val 取代。

另外一种具有更明显效应的改变是突变导致链终止密码子，称为无义突变（nonsense mutation）或者终止突变（stop mutation）。当遗传信息中出现无义突变时，由于过早的链终止，不完全的多肽链会从核糖体释放出来。不完全肽链的大小取决于无义突变的位置。如果突变靠近基因起始端，所合成的多肽链非常短；如果突变靠近基因的末端，合成的肽链长度则几乎接近正常。如第 14 章所述，如果 mRNA 包含一个过早的终止密码子，就会在真核细胞中通过一个称为无义介导的 mRNA 衰减的过程而很快地降解。

第三种点突变是移码突变（frameshift mutation）。它是插入或者缺失一个或者几个碱基对，从而改变可读框。下面以应该读成一系列 Ala 的密码子框架中的串联重复序列 GCU 为例（为了清晰起见，密码子人为地用空格分开，当然真正的 mRNA 是连续的）：

```
   Ala Ala Ala Ala Ala Ala Ala Ala
5′-GCU GCU GCU GCU GCU GCU GCU GCU-3′
```

现在假定在遗传信息中插入 1 个 A，因而在插入位点产生 1 个 Ser 密码子（AGC）。引起的移码突变使插入位点下游的三联体读成 Cys：

```
   Ala Ala Ser Cys Cys Cys Cys Cys
5′-GCU GCU AGC UGC UGC UGC UGC UGC-3′
```

因此插入或者缺失一个碱基，不但显著地改变突变位点的遗传信息的编码能力，而且同样影响下游其他的遗传信息。同样地，插入或者缺失两个碱基，其效应将扩散至突变位置下游的所有序列，导致完全不同的可读框。

最后考虑一下在遗传信息相近的位置插入 3 个额外碱基的情形。显而易见，在 3 个插入碱基位置上和之间的遗传信息会有显著的改变。但由于密码子是以 3 个核苷酸为单位的，在 3 个插入碱基的下游，mRNA 将保持正常的可读框而不发生改变。

Ala Ala Ser Cys Met Leu His Ala Ala Ala
5′-GCU GCU **A**GC UGC **A**UG CUG C**A**U GCU GCU GCU-3′

遗传密码是以 3 个碱基为阅读单位的遗传学证据

上述例子是 Francis Crick、Sydney Brenner 及其合作者用一个经典实验验证的逻辑，该实验用 T4 噬菌体建立了遗传密码以 3 个碱基为阅读单位，并且该论点的建立纯粹基于遗传学讨论（也就是说，没有生物化学和分子的证据）。遗传杂交制造出含有 3 个单碱基对插入在单个基因的相邻位点的噬菌体突变体。当然，3 个碱基插入搅乱了一小段密码子，但目标基因（称作 *r*Ⅱ）所编码的蛋白质能够忍受氨基酸序列的局部改变。这一发现表明：尽管存在 3 个变异，但基因的总体编码能力并未发生改变，这 3 个变异单独存在，或者两两存在，都会显著地改变基因信息的可读框（致使其蛋白质产物没有活性）。由于基因能够耐受 3 个碱基的插入，但不能耐受 1 个、2 个（或 4 个）碱基的插入，因此遗传密码子必须以 3 个碱基为单位进行阅读。参见第 2 章和第 22 章有关那些证明了遗传密码以 3 个碱基为单位进行阅读的历史人物的讨论，以及有关 T4 噬菌体作为模式系统在阐明遗传密码本质作用的描述。

抑制突变可以存在于相同或不同的基因中

通常，有害突变的效应可以被第 2 个遗传变化逆转。其中某些继发突变易于理解，如简单的逆转突变（reverse mutation），把改变了的核苷酸序列转变回原先的序列。比较难以理解的是突变发生在染色体的不同位置，通过在 B 位点产生其他遗传变化抑制发生在 A 位点的突变所带来的变化。这种抑制突变（suppressor mutation）主要有两种类型：一是基因内抑制（intragenic suppressor），即抑制突变和原来的突变位于同一个基因内，但在不同的位置上；二是抑制突变发生在另一个基因上的基因间抑制（intergenic suppression）。抑制其他基因的突变的基因称为抑制基因（suppressor gene）。在此我们考虑的两种突变抑制作用均通过导致产生好的（或部分好的）蛋白质拷贝而起作用，原有的有害突变则使蛋白质失活。例如，如果第一个突变使参与 Arg 合成的一个酶产生没有活性的拷贝，那么抑制突变通过恢复合成该酶某些好的拷贝，从而使 Arg 能够合成。然而，基因内抑制突变与基因间抑制突变恢复正常蛋白合成的机制是完全不同的。

以错义突变为例来说明基因内抑制。错义突变的效应有时可以通过同一个基因内的另外一个错义突变而逆转。在这种情形下，最初酶活性的丢失是由于在编码的蛋白质序列中存在一个错误的氨基酸而引起蛋白质三维构象的改变。同一个基因中的第二个错义突变可以恢复蛋白质生物学活性，如果它以某种方式恢复蛋白质分子功能集团附近的构象。图 15-6 显示了基因内抑制的另一个例子，这回是移码突变。

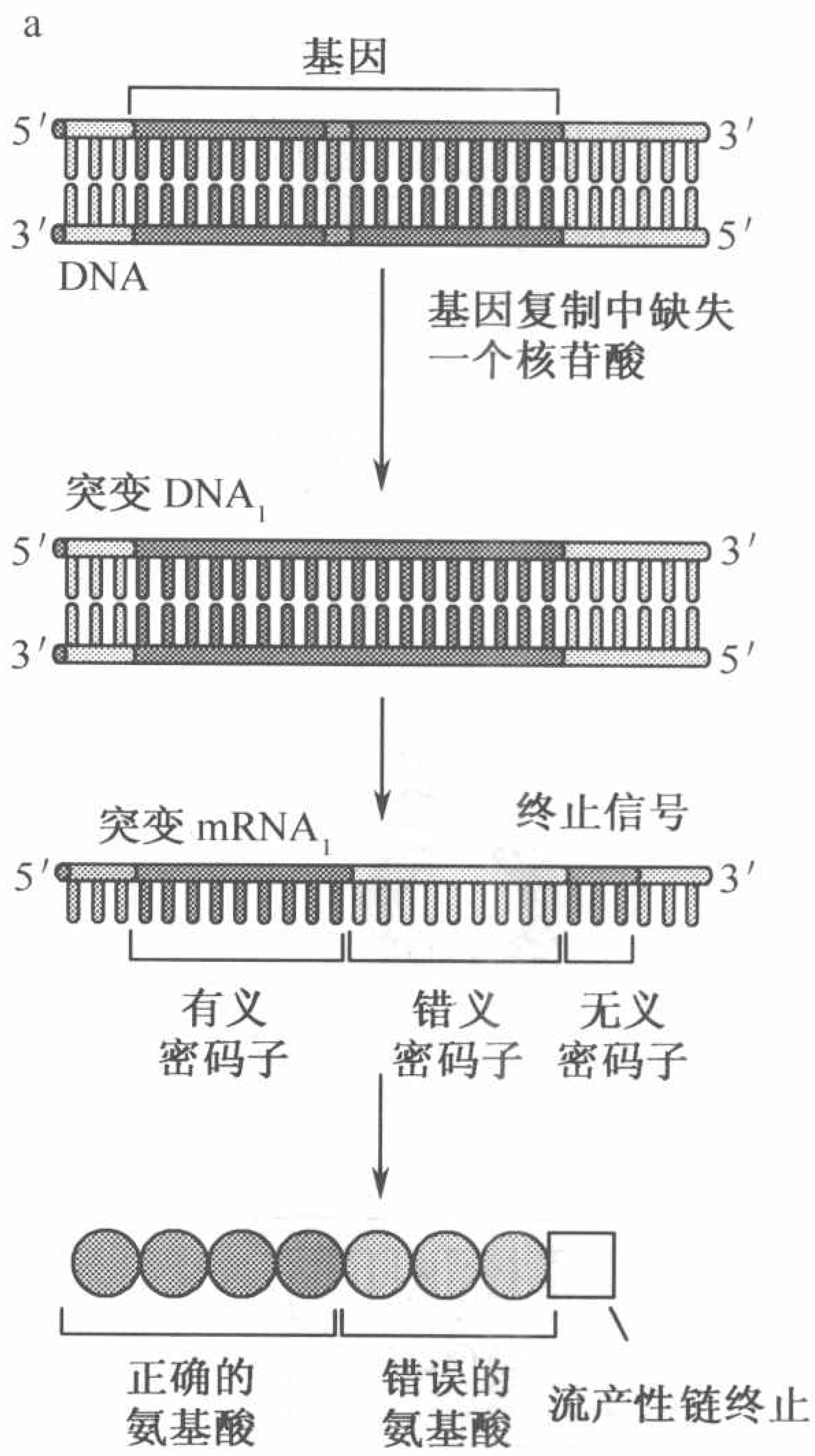


图 15-6 移码突变的抑制。 a. 核苷酸编码序列的一个缺失可以导致所合成的多肽链不完全并失去活性。b. 如 a 部分所示的缺失效应可以被编码序列的第二个突变，即一个插入所纠正。插入的纠正效果表现为合成完整的肽链，但有两个氨基酸发生替代。取决于序列中替代的不同，蛋白质可能具有部分或者全部活性。

基因间抑制涉及突变 tRNA

抑制基因并不是通过改变突变基因的核苷酸序列而发挥作用的。相反，它们改变 mRNA 模板的阅读方式。最著名的抑制突变例子之一是突变 tRNA 基因，它抑制蛋白质编码基因中无义突变产生的效应（但也有抑制错义突变，甚至移码突变的突变 tRNA)。在大肠杆菌中，3 个终止密码子都有抑制基因。它们通过把终止密码子阅读成一种特异氨基酸的信号而起作用。例如，有 3 个研究得清楚抑制 UAG 密码子的基因，在无义突变位置一个抑制基因插入 Ser，另外一个插入 Glu，第 3 个插入 Tyr。在 3 个 UAG 抑制突变中，3 种氨基酸的每一种 tRNA 的反密码子都已经发生改变。例如，Tyr 抑制子源自 $tRNA^{Tyr}$基因内的一个突变，该突变使反密码子从 GUA（3′-AUG-5′）转变为 CUA（3′-AUC-5′），从而使其能够识别 UAG 密码子（图 15-7)。Ser 和 Glu 抑制 tRNA 也源自反密码子中的单个碱基突变。

带有无义抑制子的细胞含有突变改变了的 tRNA，这一研究发现带来了这样的问题：与这些 tRNA 相应的密码子如何能继续正确地阅读呢？对 Tyr UAG 抑制子来说，答案来自于有 3 个分别独立的基因编码 $tRNA^{Tyr}$的发现。一个编码主要的 $tRNA^{Tyr}$类型。而另外两个重复的基因编码表达量比较少的那种类型。两个重复基因中的这个或那个总

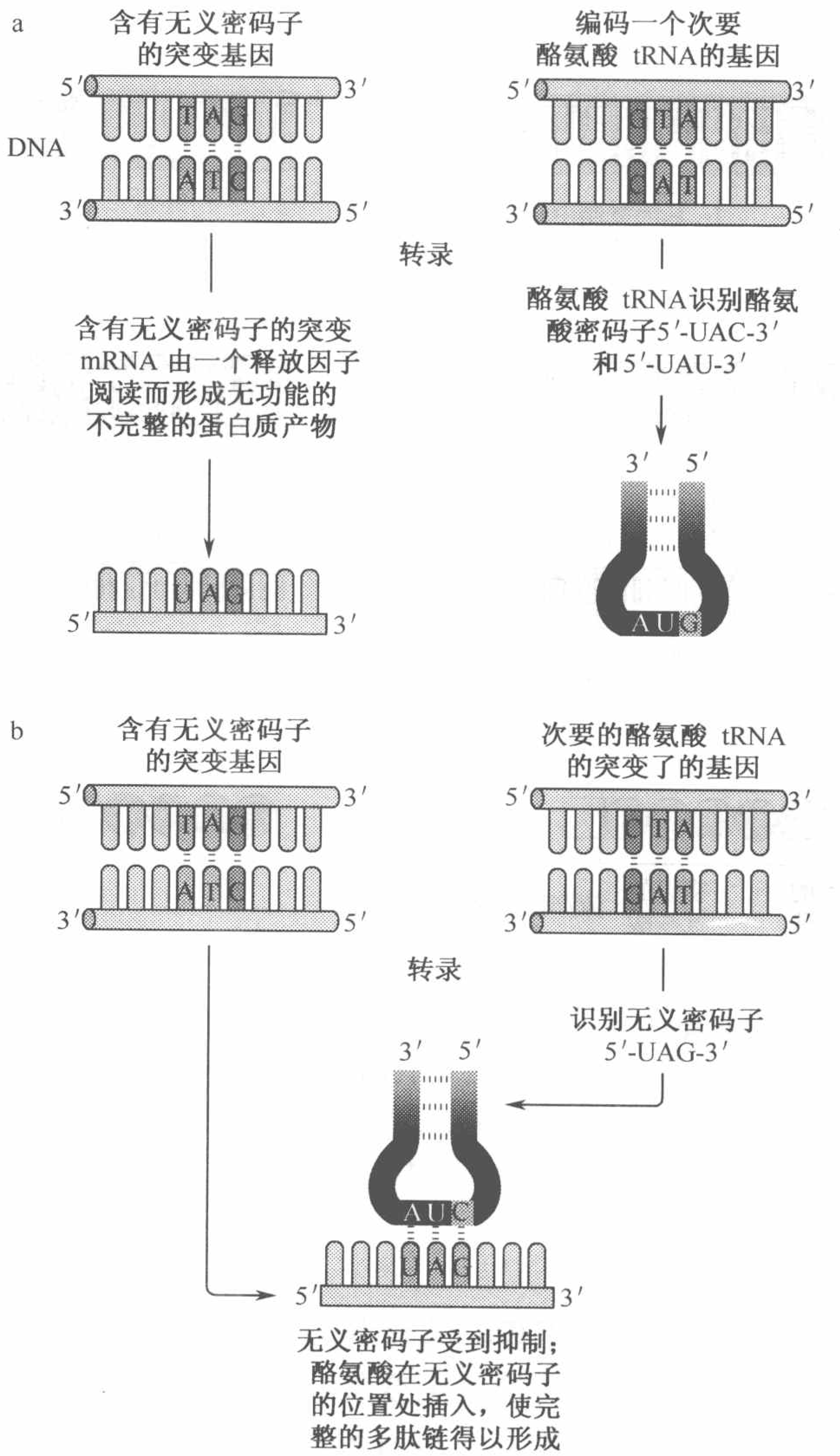


图 15-7　无义抑制。本图显示一个次要的 Tyr tRNA 如何作用以抑制 mRNA 中的无义密码子。

是抑制突变的位点。UGA 抑制就没有这种进退两难的情况，它是由一种突变形式的 $tRNA^{Trp}$介导的；抑制 $tRNA^{Trp}$仍然具有阅读 UGG（Trp）密码子的能力，同时也能识别 UGA 终止密码子。这可能是由于反密码子从 $tRNA^{Typ}$野生型的 CCA（3′-ACC-5′）转变为突变型的 UCA（3′-ACU-5′），根据前边所讨论到的摆动原理，密码子在 3′位置上的 A 或者 G 可以被反密码子 5′位置上的 U 识别。

无义突变抑制也能识别正常的终止信号

无义抑制行为可以看作是抑制 tRNA 和释放因子之间的竞争。当一个终止密码子

进入核糖体 A 位点，是读过去还是多肽链终止，取决于两者中哪一个优先到达该位置。UGA 密码子的抑制是有效的。在有抑制 tRNA 存在时，半数以上的链终止信号被读成特异的氨基酸密码子。大肠杆菌可以耐受 UAG 终止密码子的这种误读，因为 UAG 不常用作可读框末端的链终止密码子。相反，UAA 密码子抑制的平均频率通常在 1%～5%，而且产生 UAA 抑制 tRNA 的突变细胞生长缓慢。这是可以预料的，因为 UAA 常常用作链终止密码子，所以被抑制 tRNA 识别会导致产生很多异常的长肽链。

证实遗传密码的正确性

如前所述，遗传密码是采用涉及无细胞蛋白质合成系统的生物化学方法破译的。但分子生物学家一般对仅仅基于体外实验的方法持怀疑态度。那么我们如何可以确定活细胞内的遗传密码是如表 15-1 所示呢？当然，在目前大规模 DNA 测序的时代，从微生物到人类的不同生物的基因组中所有核苷酸序列已经探明，遗传密码子不但得到确证而且表明在整个生物界是通用的或几乎通用的（见下）。但是，1966 年，在人们能够进行 DNA 测序之前很久，一个经典的重要实验帮助确证了遗传密码。该实验是通过基因重组构建 T4 噬菌体的突变体，该突变体带有一对相互抑制的插入和缺失突变（和图 15-6 的实验相似）。所研究的基因编码一种细胞壁降解酶，即溶菌酶，之所以用它来进行实验，是因为该酶比较小、易于纯化，并且它的所有氨基酸序列是已知的。实验的策略是将溶菌酶双突变基因的氨基酸序列与野生型基因的氨基酸序列进行比较。

当突变的氨基酸序列（…NH_2-Thr Lys **Val His His Leu Met** Ala Ala Lys-COOH）与野生型（…NH_2-Thr Lys **Ser Pro Ser Leu Asn** Ala Ala Lys-COOH）比较时，发现其中有一串 5 个氨基酸不同（黑体所示）。这一观察结果表明插入和缺失突变已经搅乱了突变体信息中一小段密码子的排列。了解打乱的密码子对蛋白质氨基酸序列的影响为认识遗传密码本质增加了重要的约束条件。特别是，如果遗传密码如生物化学实验所揭示的那样，就可能鉴别出野生型序列 Ser Pro Ser Leu Asn 的一组密码子，在这组序列的一端插入、在另一端缺失，并经过恰如其分的排列，能够定义出突变氨基酸的序列。事实上，有办法这样做，就是要在基因编码序列的 5′端构建一个核苷酸的缺失，3′端构建一个核苷酸的插入：

NH_2-Lys	**Ser**	**Pro**	**Ser**	**Leu**	**Asn**	Ala-COOH
5′-AAA	**A**GU	CCA	UCA	CUU	AAU	GC-3′
5′-AAA	GUC	CAU	CAC	UUA	AUG	**G**C-3′
NH_2-Lys	**Val**	**His**	**His**	**Leu**	**Met**	Ala-COOH

你可以发现，这种方案可以确证几种不同的密码子，而且证明在体内不止一种的同义密码子定义相同的氨基酸（如 5′-CAU-3′和 5′-CAC-3′均定义 His）。最后一点，也是非常重要的一点，即该研究方案证明了翻译过程是沿着 5′→3′端方向进行的。（提示：当你在研究中从 3′→5′端排列每一个密码子，你能否按照从氨基端到羧基端的方向正确说明两个氨基酸的序列。）

遗传密码几乎通用

大规模基因组测序的结果已经广泛地证实了预期的遗传密码子的通用性。密码子的通用性对于我们了解进化有着重大的影响，因为它使直接比较基因组序列已知的所有生物的蛋白质编码序列成为可能。在第 21 章，我们将要讨论到，已有强有力的计算机程序，可以从大量的生物种类中搜索和发现所预测的基因编码序列的相似性。遗传密码的通用性还帮助建立了遗传工程领域，使在代理宿主生物中表达编码有用蛋白质产物的克隆基因成为可能，如在细菌中生产人类胰岛素（详见第 21 章）。

为了了解遗传密码的保守性，想一想如果遗传密码发生突变将会发生什么。例如，一个突变可能改变对应于 UCU 的那类 Ser tRNA 分子，使它转而识别 UCU 序列。这对只含有一个指导 $tRNA^{Ser}$合成的基因的单倍体细胞而言是致命的，因为 Ser 将不能插入到蛋白质序列中许多正常的位置上。即使有多于一个的 $tRNA^{Ser}$基因（如二倍体细胞），这种突变仍然是致命的，因为这将导致在细胞蛋白质中很多 Phe 被 Ser 同时取代。

从上述观点出发，对某些亚细胞器中的遗传密码和标准密码有些微差异的发现就显得完全出乎人们的意料了。这是在研究人线粒体基因组 DNA 全长 16 569bp 时才意识到的。但之前在酵母、果蝇和高等植物线粒体中已经观察到该现象。已知编码蛋白质区域的序列揭示了在标准密码子和线粒体密码子之间存在以下的差异（表 15-6）：

- UGA 编码 Trp，而非终止信号。因此线粒体 $tRNA^{Trp}$的反密码子既识别 UGG，也识别 UGA，似乎遵循传统的摆动规则。
- 非起始的 Met 既可以由 AUG 编码，也可以由 AUA 编码。
- 在哺乳动物的线粒体中，AGA 和 AGG 并不编码 Arg（在“通用”密码里 Arg 有 6 个密码子），而是指示链的终止。因此，在哺乳动物的线粒体密码中有 4 个终止密码子（UAA、UAG、AGA 和 AGG）。
- 在果蝇的线粒体中，AGA 和 AGG 也不编码 Arg，而是编码 Ser。

也许并不奇怪，谈到解码线粒体信息的规则，线粒体 tRNA 同样不同寻常。哺乳动物线粒体中只有 22 个 tRNA，但根据摆动原则，至少需要 32 种 tRNA 分子来解码通用密码子。结果，当一个氨基酸有 4 种密码子（密码子第 1 和第 2 位置的核苷酸相同）来定义时，只涉及单一的线粒体 tRNA。（而在非线粒体系统，至少需要 2 种 tRNA 分子。）线粒体 tRNA 的反密码子 5′（摆动）位置上的核苷酸都是一个 U，它可以和密码子 3′位置上的 4 种核苷酸中的任何一种进行配对。如果密码子 3′位置上是由嘌呤碱基组成的密码子，与由嘧啶碱基组成的密码子编码不同的氨基酸，那么线粒体 tRNA 反密码子 5′第一核苷酸位置上是经过修饰的 U，它限制摆动，只与两个嘌呤配对。

通用密码子例外的情况不限于线粒体，也见于几个原核生物基因组和某些真核生物的核基因组。细菌山羊支原体（*Mycoplasma capricolum*）用 UGA 编码 Trp，而非终止密码子。同样地，一些单细胞原生动物用 UAA 和 UAG 编码 Gln，而在“通用”密码子中，它们是终止密码子。最后，在“通用”密码中一个氨基酸（Leu）的密码子

表 15-6 哺乳动物线粒体的遗传密码*

第 1 个位置（5′端）	第 2 个位置 U	第 2 个位置 C	第 2 个位置 A	第 2 个位置 G	第 3 个位置（3′端）
U	UUU Phe UUC (GAA)† UUA Leu UUG (UAA)	UCU UCC Ser UCA (UGA) UCG	UAU Tyr UAC (GUA) UAA 终止子 UAG 终止子	UGU Cys UGC (GCA) UGA Trp UGG (UCA)	U C A G
C	CUU CUC Leu CUA (UAG) CUG	CCU CCC Prn CCA (UGG) CCG	CAU His CAC (GUG) CAA Gln CAG (UUG)	CGU CGC Arg CGA (UCG) CGG	U C A G
A	AUU Ile AUC (GAU) AUA Met AUG (GAU)‡	ACU ACC Thr ACA (UGU) ACG	AAU Asn AAC (GUU) AAA Lys AAG (UUU)	AGU Ser AGC (GUC) AGA 终止子 AGG 终止子	U C A G
G	GUU GUC Val GUA (UAC) GUG	GCU GCC Ala GCA (UGC) GCG	GAU Asp GAC (GUC) GAA Glu GAG (UUC)	GGU GGC Gly GGA (GCC) GGG	U C A G

* 线粒体遗传密码与“通用”遗传密码（表 15-1）的区别用阴影表示。

† 每一组密码子用阴影表示并有单一的 tRNA 来阅读，该 tRNA 的反密码子以 5′→3′写在括号内。每 4-密码子组被 1 个 tRNA 阅读，该 tRNA 的反密码子的第一个（5′）位置有一个 U。末端为 U/C 或者 A/G 的 2-密码子组在通过 tRNA 的 GU 摆动来阅读，反密码子第一位置分别是 G 或 U。反义密码子经常含有修饰碱基。

‡ 注意在反义密码子第一位置的 C 进行不寻常的配对。

(CUG)，在假丝酵母（*Candida*）中变成了另一个氨基酸（Ser）的密码子。

小结

从细菌到人类每一种生物都“通用”的遗传密码中，61 种密码子编码特定的氨基酸；其余 3 种是终止密码子。密码子有高度的简并性，即几种密码子（同义）通常与一种氨基酸相对应。有时一个特定的 tRNA 可以特异地识别几种密码子，这种能力来自于反密码子 5′端碱基的摆动。终止密码子 UAA、UAG 和 UGA 由特异的蛋白质识别，而非特异的 tRNA 分子。

遗传密码子遵守 3 个原则：密码子从 5′→3′方向进行阅读；密码子相互不重叠并且信息没有缝隙；信息是在固定的可读框里翻译的，可读框由起始密码子设定。

遗传密码是通过在无细胞提取物中研究蛋白质合成而破译的。在原信使组分消耗殆尽的提取物中加入新的 mRNA 导致产生新的蛋白质，它们的氨基酸序列是由外加的

mRNA 所决定的。对合成的多聚核糖核酸 poly-U 可以特异地编码 poly-Phe 的发现是遗传密码破译的第一步（也许是最重要的一步）。用其他合成的均一的多聚核糖核酸（如 poly-C 等）或者混合的多聚核糖核酸（如 poly-AU 等），使不同氨基酸的密码子组成得以确定。密码子中核苷酸的准确排列的最终确定则源自随后的对于特异的三核苷酸-tRNA-核糖体相互作用的研究以及常见的共聚物作为信使的使用。

改变密码的点突变包括错义突变，它使一个氨基酸的密码子变为另外一个氨基酸的密码子；无义突变可以使蛋白质的合成过早终止；移码突变则改变遗传信息的可读框。在一些情况下，错义突变、无义突变和移码突变的效应可以部分地被基因外抑制所逆转。例如，突变的 tRNA 可以将由于无义突变造成的终止密码子阅读成编码特异的氨基酸的密码子。

在线粒体基因组及一些原核生物、单细胞原生动物的主要基因组中使用些稍微不同的遗传密码，如“通用密码”中的终止密码子 UGA，在线粒体、原核生物、原生动物基因组中编码 Trp。

（闫春霞 译 吴琳 校）

参考文献

书

Celis J.E. and Smith J.D., eds. 1979. *Nonsense mutations and tRNA suppressors.* Academic Press, New York.

Clark B. and Petersen H., eds. 1984. Gene expression: The translational step and its control. *Alfred Benzon Symposium,* vol. 19. Copenhagen, Munksgaard.

Cold Spring Harbor Symposia on Quantitative Biology. 1966. Volume 31: *The genetic code.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Söll D.G., Abelson J.N., and Schimmel P.R., eds. 1980. *Transfer RNA: Biological aspects.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Ycas M. 1969. *The biological code.* Wiley (Interscience), New York.

遗传密码的特征

Crick F.H.C. 1966. Codon–anticodon pairing: The wobble hypothesis. *J. Mol. Biol.* **19:** 548–555.

Kohli J. and Grosjean H. 1981. Usage of the three termination codons: Compilation and analysis of the known eukaryotic and prokaryotic translation termination sequences. *Mol. Gen. Genet.* **182:** 430–439.

Lagerkvist U. 1981. Unorthodox codon reading and the evolution of the genetic code. *Cell* **23:** 305–306.

密码是如何被打断的

Crick F.H.C. 1963. The recent excitement in the coding problem. *Prog. Nucleic Acid Res.* **1:** 164.

Khorana H.G. 1968. *Polynucleotide synthesis and the genetic code. Harvey Lecture Series 1966–67.* Vol. 62. Academic Press, New York.

Nirenberg M. and Leder P. 1964. The effect of trinucleotides upon the binding of sRNA to ribosomes. *Science* **145:** 1399–1407.

Speyer J.F., Lengyel P., Basilio C., Wahba A.J., Gardner R.S., and Ochoa S. 1963. Synthetic polynucleotides and the amino acid code. *Cold Spring Harbor Symp. Quant. Biol.* **28:** 559–568.

遗传密码遵循的三条规律

Brenner S., Stretton A.O.W., and Kaplan S. 1965. Genetic code: The nonsense triplets for chain termination and their suppression. *Nature* **206:** 994–998.

Crick F.H.C., Barnett L., Brenner S., and Watts-Tobin R.J. 1961. General nature of the genetic code for proteins. *Nature* **192:** 1227–1232.

Garen A. 1968. Sense and nonsense in the genetic code. *Science* **160:** 149–159.

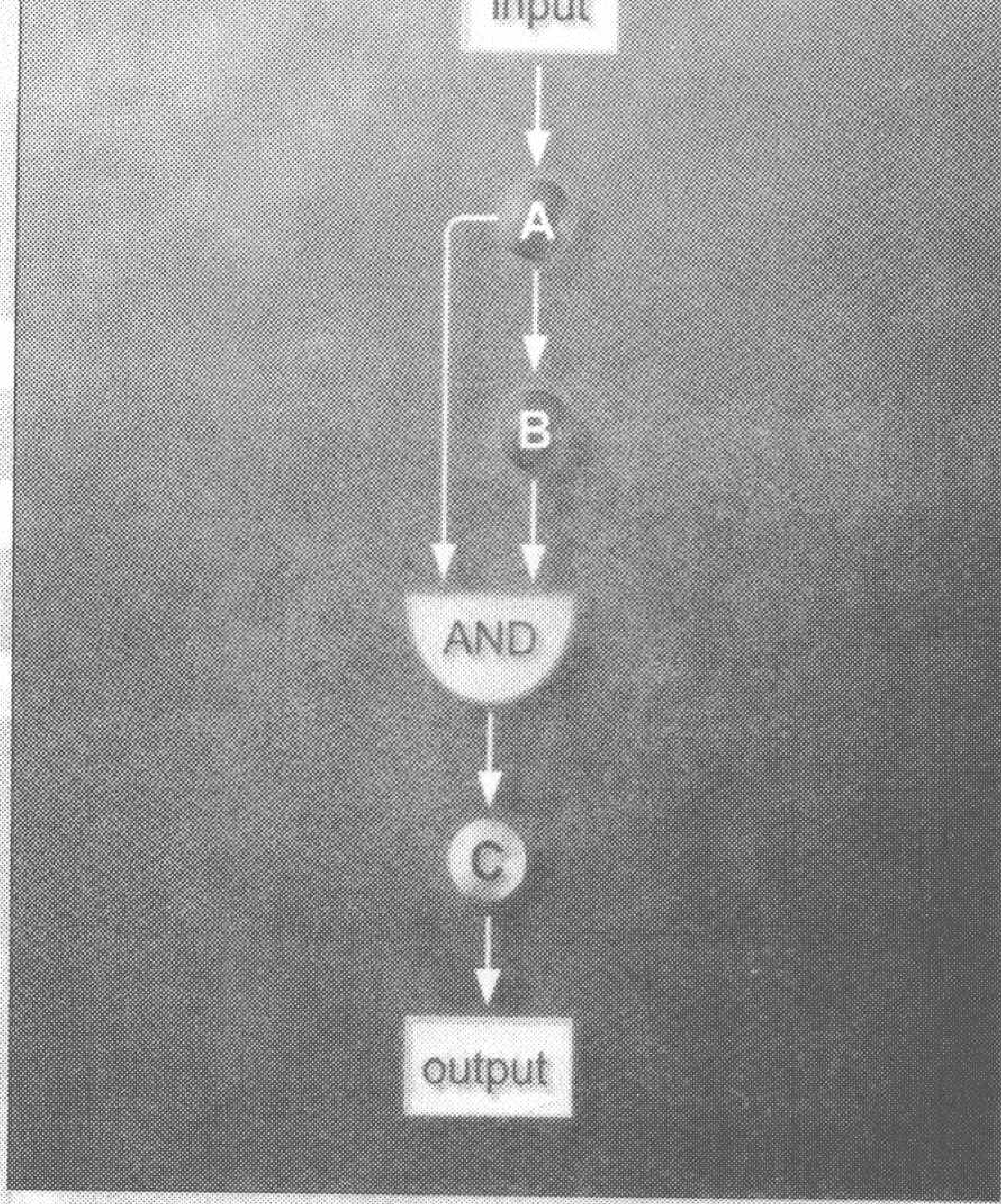
Terzaghi E., Okada Y., Streisinger G., Emrich J., Inouye M., and Tsugita A. 1966. Change of a sequence of amino acids in phage T4 lysozyme by acridine-induced mutations. *Proc. Natl. Acad. Sci.* **56:** 500–507.

抑制

Buckingham R.H. and Kurland C.G. 1980. Interactions between UGA-suppressor tRNA'P and the ribosome: Mechanisms of tRNA selection. In *Transfer RNA: Biological aspects* (ed. D. Söll et al.), pp. 421–426. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Ozeki H., Inokuchi H., Yamao F., Kodaira M., Sakano H., Ikemura T., and Shimura Y. 1980. Genetics of nonsense suppressor of tRNAs in *Escherichia coli.* In *Transfer RNA: Biological aspects* (ed. D. Söll et al.), pp. 341–349. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Steege D.A. and Söll D.G. 1979. Suppression. In *Biological regulation and development I* (ed. R.F. Goldberger), pp. 433–486. Plenum, New York.



第4篇

调　　控

本篇概要

- **第 16 章 原核生物的基因调控**
- **第 17 章 真核生物的基因调控**
- **第 18 章 调控 RNA**
- **第 19 章 发育和进化的基因调控**
- **第 20 章 基因组分析与系统生物学**

在第 3 篇，我们讲述了 DNA 编码的遗传信息是如何表达的，包括从 DNA 序列转录为 RNA 和以 RNA 为模板翻译为蛋白质。

并非所有细胞的全部基因总在不断地表达。实际上，生命很大程度上依赖于细胞在不同时间、不同空间以不同的组合、表达不同的基因。甚至一种低等的细菌在特定时间里也只表达某些基因——保证可以合成某些酶以代谢它所遇到的营养物，同时却不合成其他营养物代谢所需的酶。多细胞有机体的发育为这种所谓的“差异基因表达”提供了典型的例子。基本上，所有的人类细胞包含相同的基因，但是在形成某种类型的细胞时表达的一套基因与形成另一类型的细胞时所表达的基因是不同的。因此，肌肉细胞中表达的一套基因（至少是部分的）与神经元或者皮肤细胞等所表达的基因是不同的，这些不同出现在转录水平上——最常见于转录起始上。

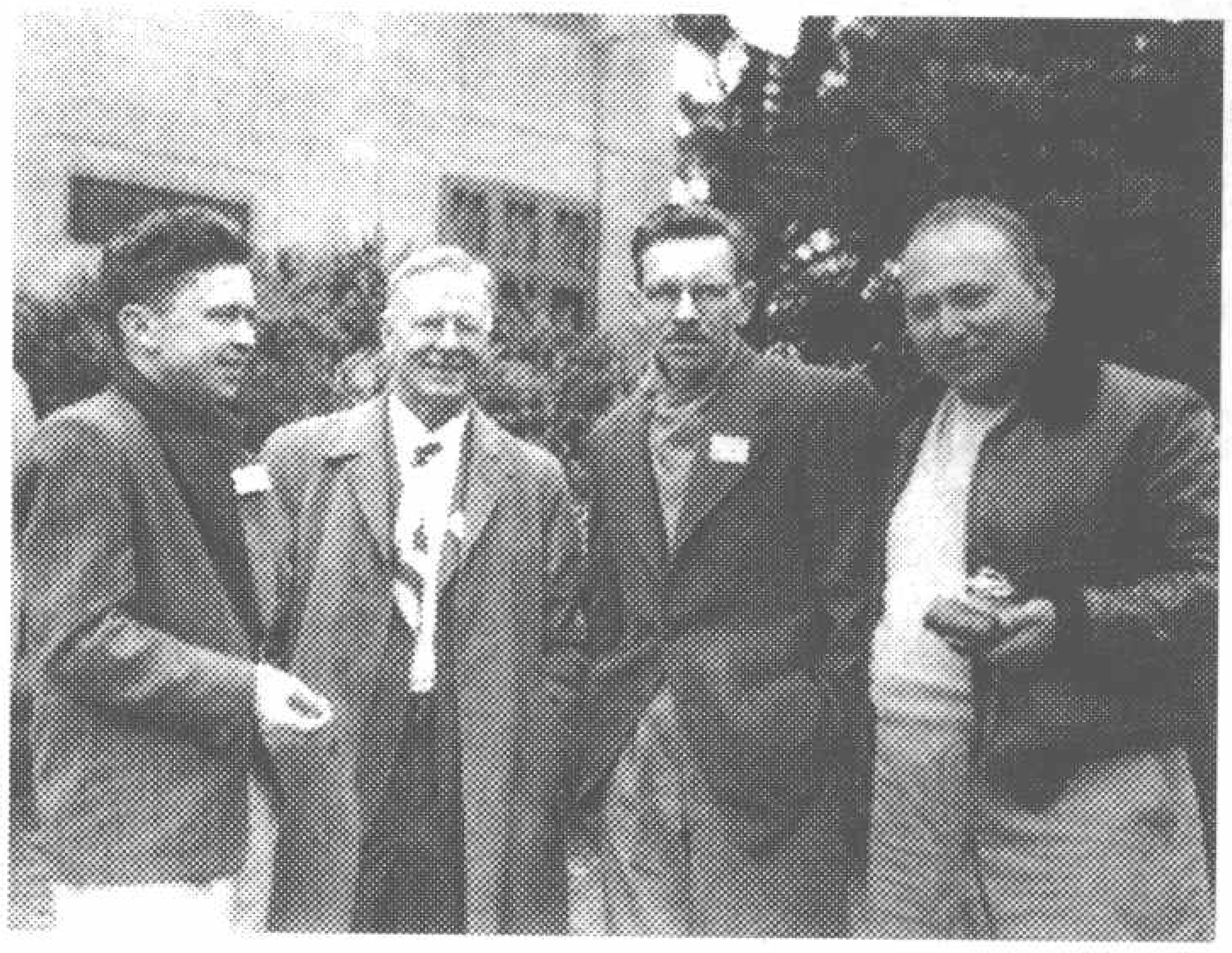
在接下来的章节中，我们将看一下基因是如何被调控的。第 16 章将讲述细菌中的基因调控，正是从细菌体系中得到了最基本的调控原理。首先，将举几个简单的例子展示转录调控的不同原理，其中包括 *lac* 操纵子，它们的基因编码代谢乳糖所需的蛋白质，并只在培养基中有乳糖时才表达；接着将讨论操纵基因表达的后续步骤中的基因调控，例如，RNA 延伸和翻译；最后，我们会描述噬菌体 λ 在感染细菌时如何通过表达不同的基因来选择发育途径。

在第 17 章中，我们将讨论真核生物基因表达的基本机制，包括酵母和高等真核生物中发现的简单实例，并将转录激活和抑制的机制与细菌的进行比较，可以看到哪些机制是保守的，哪些机制是新增的——最值得指出的是染色质修饰的作用（第 7 章讨论的类型）。我们也将看到小 RNA 分子如何以各种方式调控基因表达。

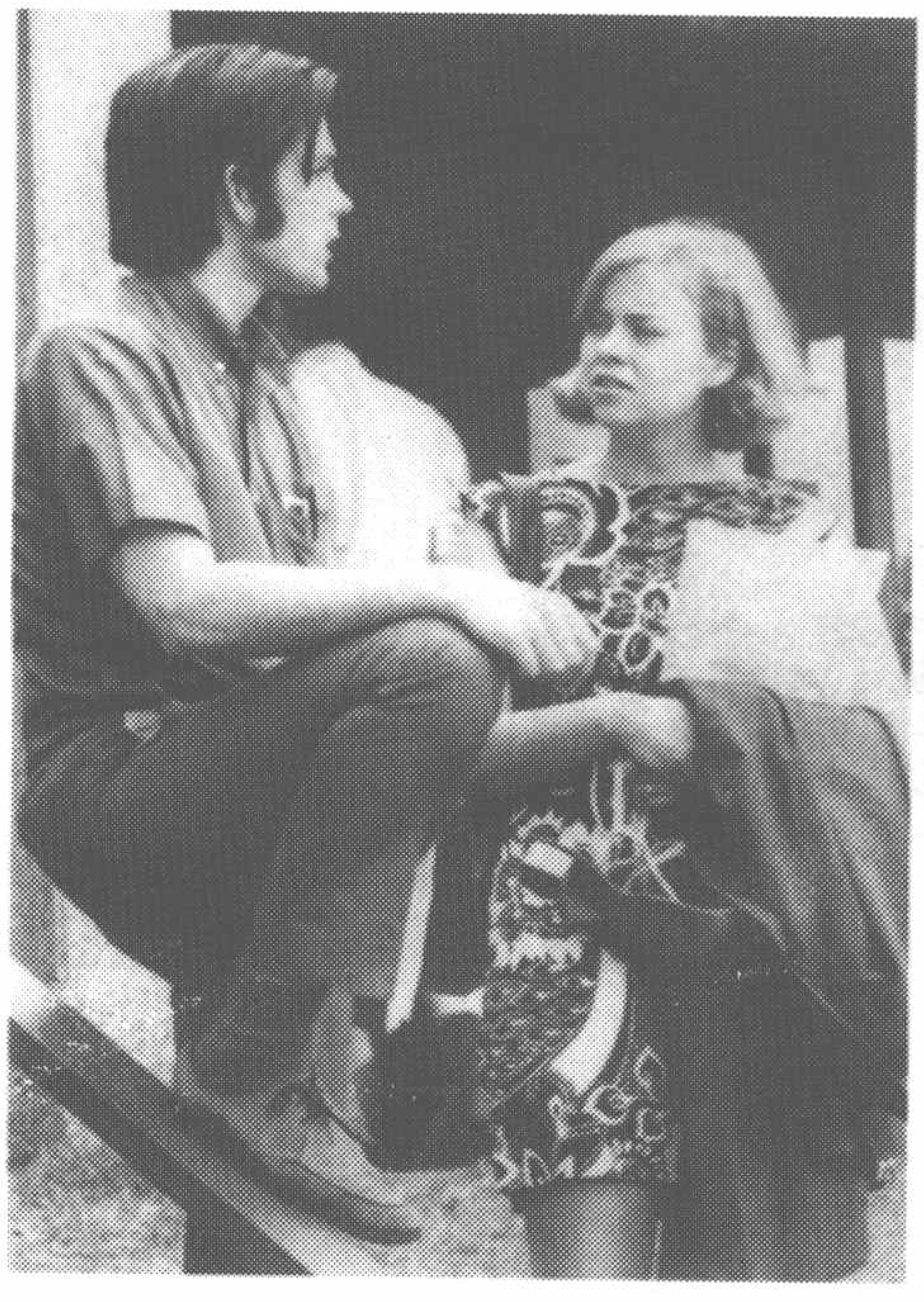
第 13 章曾讲到，在真核生物中，RNA 的剪接经常发生在翻译之前，这为一种特定基因的调控提供了另一个步骤。在这种情况中，调控不仅能决定某一基因何时表达，还可以决定合成几个选择性蛋白质中的哪一个。

第 18 章和第 19 章将从发育生物学的角度中讨论基因调控。在第 18 章，我们会讲述在一群遗传上等同的细胞（如发育中的胚胎）中基因如何调控以赋予细胞种类特异性（分化）和类型形成（形态发生）。第 19 章将帮我们去理解亲缘关系很近的生物的多样性以及其中许多生物的形态和行为差异不是由于基因的变化造成的，而是由于每种生物中的基因在发育过程中表达的位置及时间的不同导致的。

冷泉港实验室资料库中的照片



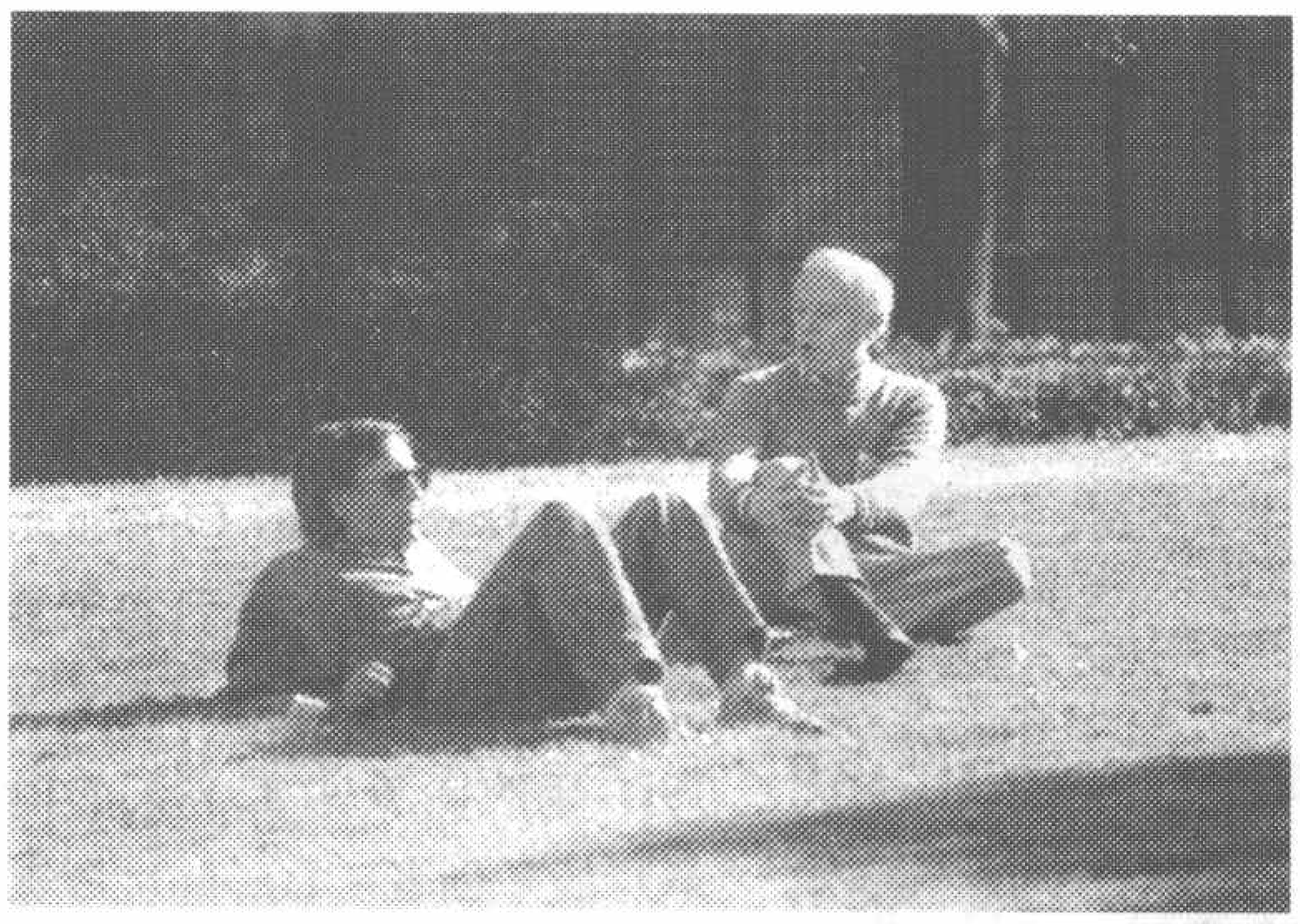
Edward Lewis，Carl Lindegren，Alfred Hershey 和 Joshua Lederberg　1951 年，基因和突变讨论会。Lewis 用果蝇作为模型（第 18 章）倡导了发育的遗传学分析，并因此分享了 1965 年的诺贝尔医学奖。Lindegren 是酵母遗传学（第 21 章）的开创者。Hershey 带领包括 Max Dedlbruk 和 Salvador Luria 的研究小组在分子生物学的早期应用噬菌体作为模型系统（第 21 章），他们三人共同获得 1969 年的诺贝尔医学奖。Lederberg 则发现了 DNA 在细菌间可以通过被称为接合的交配过程传递（第 21 章），他也因此而分享了 1958 年的诺贝尔医学奖。



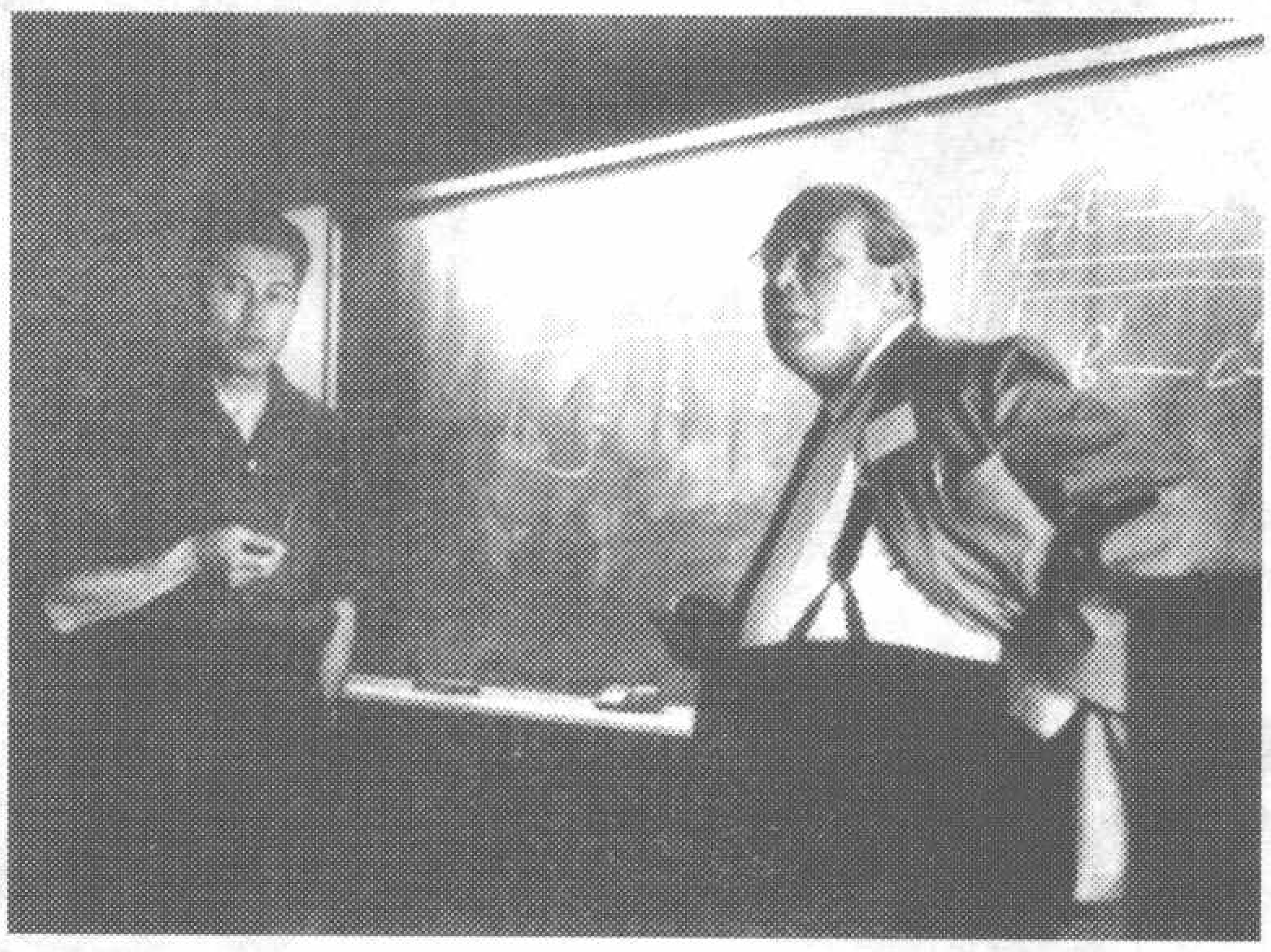
Jeffrey Roberts 和 Ann Burgess　1970 年，遗传物质转录讨论会。Roberts 的研究集中在细菌和噬菌体中调控基因表达的调控蛋白，特别是 lambda 中的抗终止子（第 16 章）。Burgess 是一名生物学教育者，并参与全国性的改进科学教育活动。Roberts 是本书前一版的作者之一，而本版作者之一 Tania Baker（TB）是 Burgess 的亲戚。



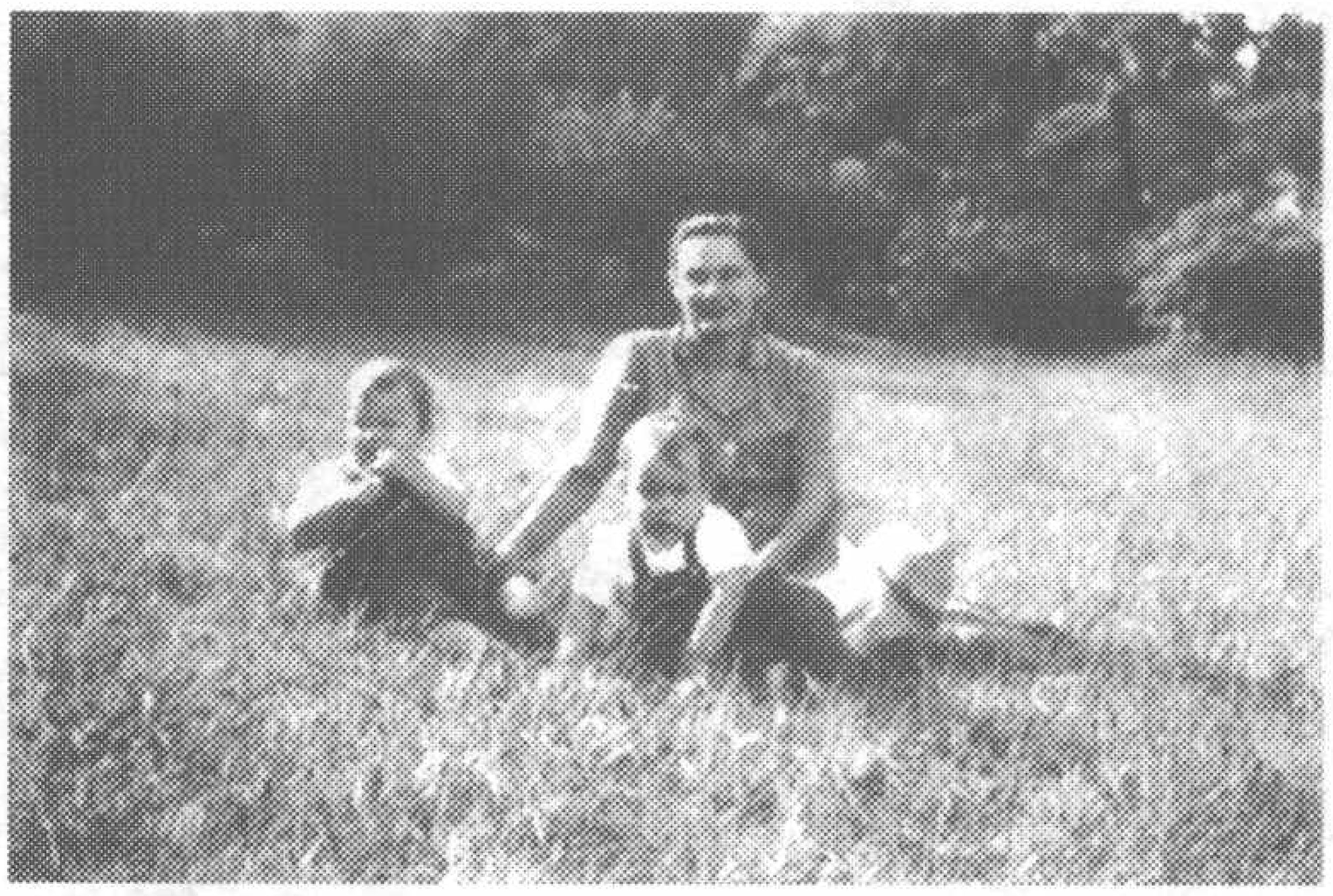
Christiane Nüsslein-Volhard　1996 年，冷泉港实验室斑马鱼的发育和遗传学会议。由 Nusslein-Volhard 和她的同事 Eric Wieschaus 在果蝇中进行的突变筛选确认了果蝇（可能也是所有动物）的早期胚胎发育中的许多关键基因（第 18 章），两人也因此与 Edward Lewis 共享了 1995 年的诺贝尔奖。



Mark Ptashne 和 Joseph Goldstein 1988 年，分子生物学和信号转导讨论会。受到 Jacob 和 Monod 关于基因表达调控的早期观点启发，Ptashne 描述了这些机制在分子水平上是如何起作用的（第 16 和 17 章）。Goldstein 和其长期的合作伙伴 Michael Brown 发现了控制参与胆固醇代谢的基因表达的信号转导路径（第 17 章），他们因此获得 1985 年的诺贝尔医学奖。



Jacques Monod 和 Leo Szilard 1961 年，冷泉港实验室。Monod 与 Francois Jacob 一起阐明了基因表达调控的操纵子模型（第 16 章），他们与同事 Andre Lwoff 因这一成就获得了 1963 年的诺贝尔医学奖。Leo Szilard 是战时核物理学家，在参加了 1947 年冷泉港的抗菌素课程后转向了分子生物学，他和 Aaron Novick 在芝加哥主持一个实验室（来源：Courtesy of Esther Bubley）。



I. H. Herskowitz 夫人与儿子 Ira 和 Joel 在一起 1947 年，核酸和核蛋白讨论会。Ira Herskowitz 首创了用酿酒酵母（*S. cerevisiae*）作为分子生物学研究的模式生物（第21章），像他早期研究噬菌体 *lambda* 一样，他对酵母中基因调控的观点也做出了主要贡献。其父亲 *Irwin*，后来是一本遗传学教科书的作者，当年正参加讨论会。



Craig Mello，2004 年表观遗传学讨论会。与 Andrew Fire 一起，Mello 发现将 dsRNA 导入细胞，与该 RNA 同源的基因可以被沉默。正是这项被他们成为 RNA 干扰的发现，开拓了整个 RNAi 领域（第 18 章）。他们分享了 2006 年诺贝尔生理学或医学奖。



Scott Emmons，Gary Ruvkun 和 Barbara Meyer，2004 年表观遗传学讨论会。在研究蠕虫的发育遗传学过程中，Victor Ambros 和 Ruvkun 鉴定了第一个 miRNA 和它的靶基因（第 18 章）。美国国家航空航天局（NASA）的 T 恤暗示了 Ruvkun 其他方面的兴趣：在火星上寻找生命。Emmons 在从基因表达的神经生物学的各个水平研究蠕虫的行为，Meyer 作为对阐明 λ 噬菌体遗传转换有重要贡献的研究生，目前研究蠕虫的性别决定和剂量补偿（第 18 章）。



Richard Jorgensen 和 David Baulcombe，2006 年调控 RNA 讨论会。Jorgensen 发现牵牛花色素基因的过表达会产生白色花而不是深紫色花（第 18 章）。虽然当时并不清楚产生这种现象的原因，但现在知道是由于 RNAi。小干扰 RNA——该过程的重要中介物——后来被 Baulcombe 鉴定（第 18 章）。

第16章 原核生物的基因调控

本章概要

•

转录调控的原理

•

转录起始的调控：原核生物的实例

•

λ噬菌体：调控的层次

在第12章我们了解了DNA是怎样由RNA聚合酶转录成RNA的，同时我们还描述了启动子（位于基因起点、结合转录酶和起始转录的区域）的组成组分。细菌中最常见的RNA聚合酶（带有σ^{70}）识别的启动子由3种组分组成，即“－10”、“－35”和“上游”组分。我们已知某一特定启动子的活性强度决定于其拥有上面3种组分中的哪些及其在多大程度上达到最佳匹配。在缺乏调控因子的情况下，这些组分决定了聚合酶的结合效率，一旦有了调控因子，它们便主导了转录的起始。

本章我们将看一下表达调控的机制，即细胞根据对基因产物需求的变化提升或降低表达的机制。基因的表达可在多个阶段受到调控，最为常见的是在转录起始阶段。本章的大量篇幅将集中在细菌的相应调控步骤上。我们将从调控的一般机制和原则开始，接着看几个研究较为透彻的例子，这几个例子将展示基本的调控机制是怎样以不同组合方式控制基因在特定的生物背景中的表达。我们也会论及转录后运作的基因调控机制，这包括转录抗终止作用和翻译调控。

转录调控的原理

基因表达由调控蛋白控制

正如我们在这一部分的序言中所描述的，基因常常由外部信号控制。在细菌中，主要是由培养基中存在的分子信号控制。这些信号由调控蛋白传送给基因。调控蛋白可分为两类：正调控蛋白或活化子（activator，又译激活物），负调控蛋白或抑制子（repressor，又译阻遏物）。这些调控蛋白通常都是DNA结合蛋白，它们识别受其控制的基因上的或基因附近的特异位点。活化子增强受调控基因的转录，抑制子降低或消除相应基因的转录。

这些调控蛋白是如何作用的呢？我们来回想一下第12章中描述的转录起始步骤（图12-3）。首先，RNA聚合酶在封闭复合体（closed complex）（其中的DNA双链仍在一起）中结合到启动子上。接着，聚合酶-启动子复合体转变为开放复合体，这其中

转录起始位点的 DNA 是解链的，此时聚合酶定位以便起始转录。接下来是启动子脱逃（promoter escape），这时聚合酶离开启动子开始转录。活化子在哪一步激活、抑制子在哪一步抑制，这要看具体的启动子和调控蛋白而定。接下来，我们将看到两种类型的调控。

很多启动子通过协助 RNA 聚合酶结合 DNA 的活化子和阻碍两者结合的抑制子进行调控

没有调控蛋白时，RNA 聚合酶只是微弱地结合在启动子上。这是因为该启动子缺少一种或多种上文论及的启动子组分，或这些启动子组分不完整。当聚合酶结合了启动子时，它会自动地经由封闭复合体到开放复合体的转变从而启动转录。这导致基因的组成型表达（constitutive expression），或称为本底水平（basal level）表达。在这一情况中，RNA 聚合酶的结合是限速步骤（图 16-1a）。

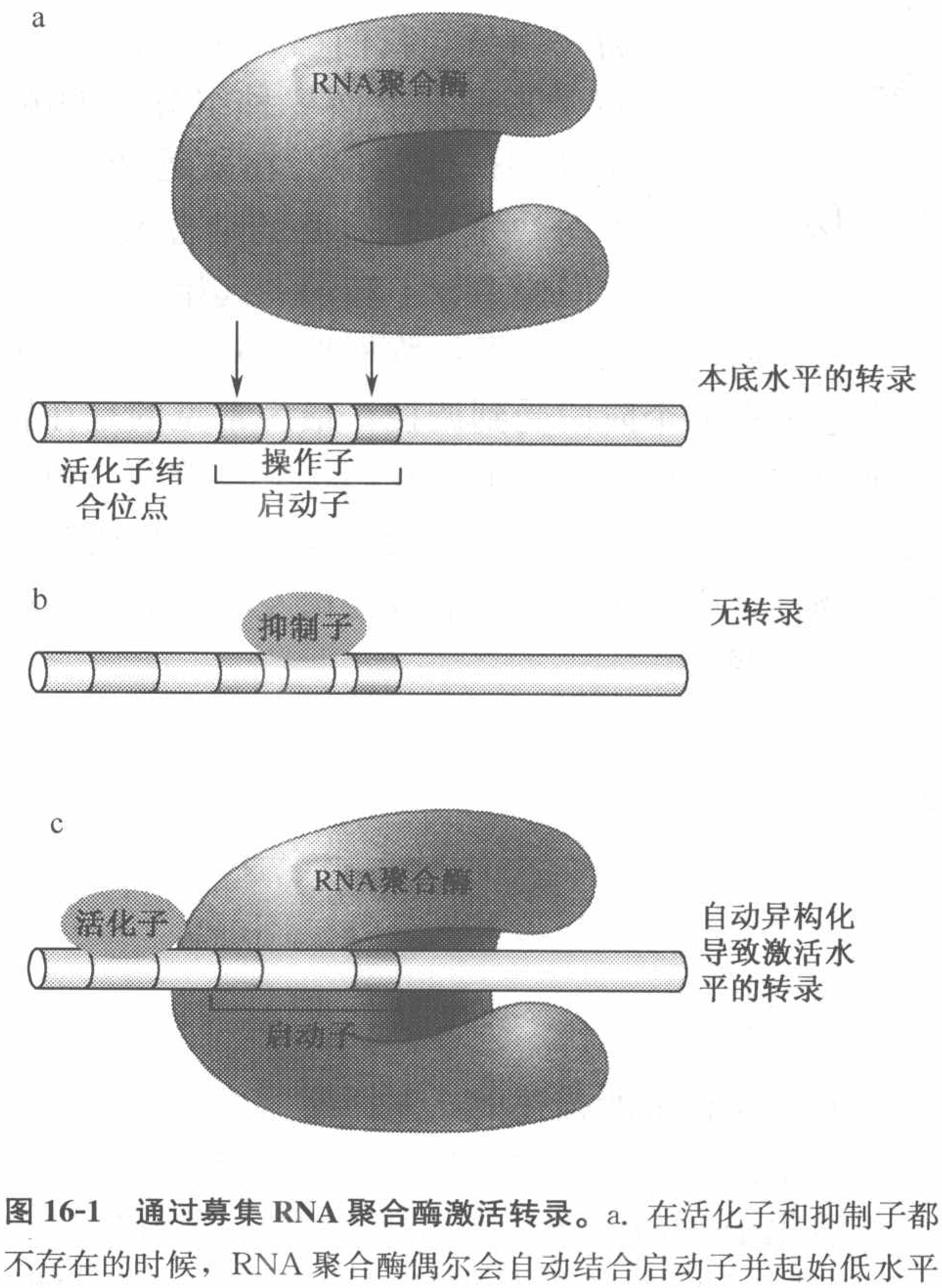


图 16-1 通过募集 RNA 聚合酶激活转录。 a. 在活化子和抑制子都不存在的时候，RNA 聚合酶偶尔会自动结合启动子并起始低水平（基本水平）的转录。b. 抑制子结合了操作子序列阻碍了 RNA 聚合酶的结合从而抑制了转录。c. 活化子募集 RNA 聚合酶造成了高水平的转录。RNA 聚合酶被募集形成封闭复合体，接着，封闭复合体自动异构化为开放复合体并起始转录。如果抑制子和活化子都存在并都有功能，一般来说，抑制子活性会优于活化子（这种情况在本图中未展示）。

抑制子要控制从某一启动子起始的表达，只需结合到与聚合酶结合区相重叠的位点上。抑制子就以此阻碍聚合酶结合到启动子上从而阻止转录（图 16-1b）。当然，抑制也可以其他方式实现，这一点也很重要。DNA 上抑制子结合的位点称为操作子（operator）。

活化子通过协助聚合酶结合到启动子上从而激活从该启动子开始的转录。通常，这是靠如下步骤实现的：活化子以其一个表面结合到启动子附近的某一 DNA 位点；同时，以另一表面与 RNA 聚合酶相互作用，将聚合酶带到启动子（图 16-1c）。这一机制常被称为募集（recruitment），这是一个蛋白质协同结合 DNA（cooperative binding of proteins to DNA，第 5 章）的例子。活化子和聚合酶的相互作用以及活化子与 DNA 的相互作用只起“黏合”作用：酶是有活性的，而活化子只是将酶带到启动子附近。一旦聚合酶结合到启动子，它就自动地使封闭复合体异构化为开放复合体并起始转录。

E. coli 的 *lac* 基因就是由活化子和抑制子以这种简单的方式调控转录的。本章我们还将详细讲述这一基因的转录调控。

某些活化子通过变构和调控 RNA 聚合酶对结合后的转录步骤起作用

启动子的调控方式并不局限于一种。因此，我们需要从上面提到的简单转录调控方式的反面来考虑。在这种情况下，RNA 聚合酶不需协助就结合在 DNA 上并形成稳定的封闭复合体，但是这一封闭复合体却不会自动转变成开放复合体（图 16-2a）。在这一启动子上，活化子必须刺激闭合复合体转变为开放复合体，因此这一转变就是限速步骤。

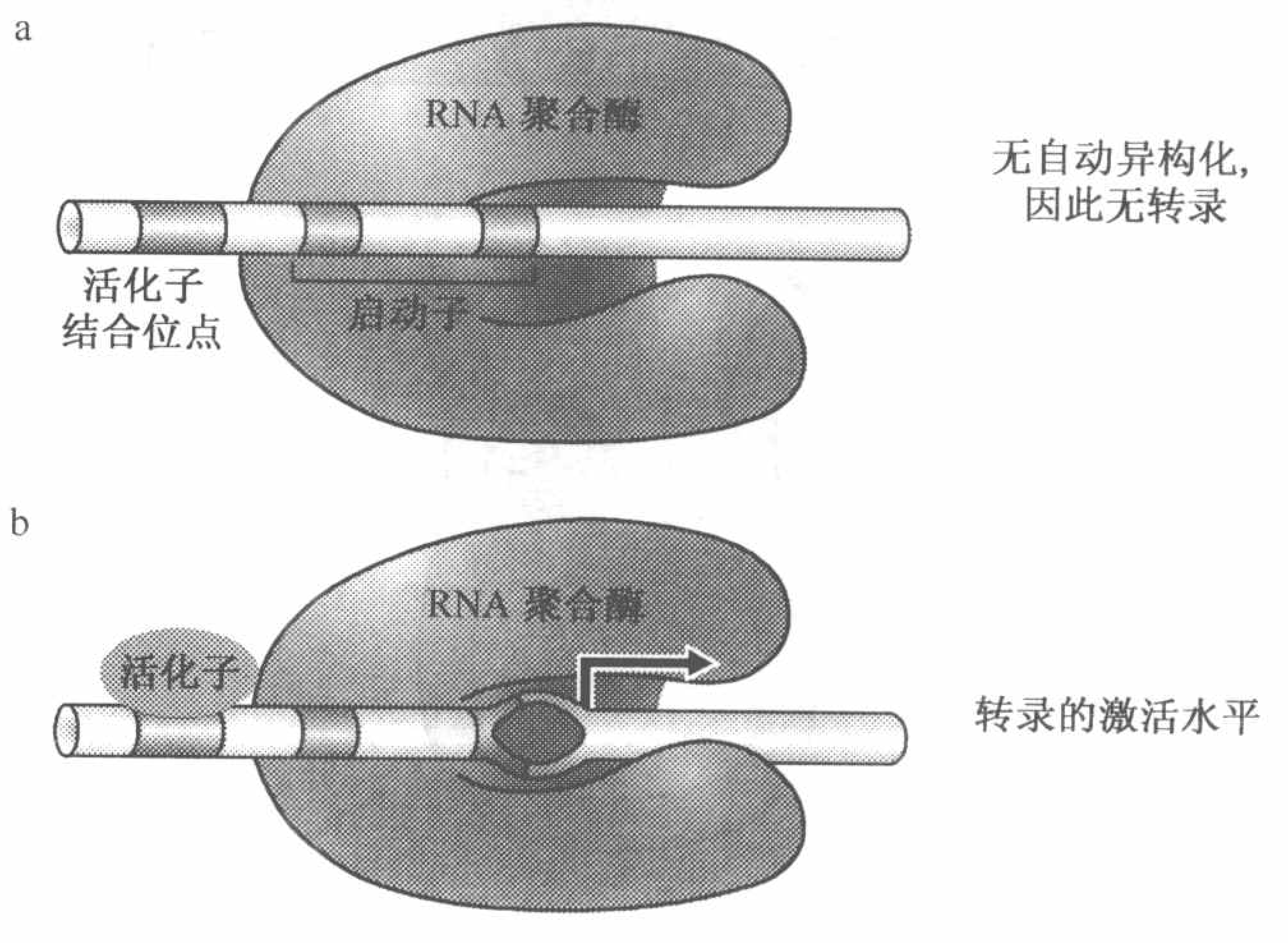


图 16-2 变构激活 RNA 聚合酶。 a. RNA 聚合酶结合封闭复合体中的启动子；b. 活化子和聚合酶相互作用引起转变为开放复合体及高水平转录。该图仅以图解的形式展示复合物的开关情况，详细解释请参照第 12 章。

刺激这类启动子的活化子通过引起 RNA 聚合酶或 DNA 的构象改变起作用。也就是说，它们与稳定的封闭复合体相互作用诱导发生构象改变引起封闭到开放复合体的转变（图 16-2b）。这一机制就是变构（allostery）的一个例证。

在第 5 章我们遇到过的变构是一种常见的控制蛋白质活性的机制。当时我们考虑过这样一个例子，蛋白质（细胞分裂素）结合并激活一个激酶（Cdk）从而在细胞分裂调控中起作用。细胞分裂素通过诱导激酶构象改变，将它从非活性状态改变为活性状态（图 5-27）。本章中，我们将看到两个转录活化子通过变构起作用的例子。在第一例（*glnA* 启动子）中，活化子（NtrC）与在启动子上结合的封闭复合体中的 RNA 聚合酶相互作用，刺激转变为开放复合体。在另一例（*merT* 启动子）中，活化子（MerR）却是通过诱导启动子 DNA 的构象改变达成同一效应的。

关于这一主题也有例外：有些启动子在多种步骤中都是失活的，同时也能够被多种机制激活。而抑制子也可以多种方式作用而不仅仅是阻碍 RNA 聚合酶的结合。例如，有些抑制子抑制封闭到开放复合体的转变或者启动子脱逃。在本章后面的内容中我们将看一些这方面的例子。

远程激活和 DNA 环化

到目前为止我们已经假定了 DNA 结合蛋白彼此相互作用结合到临近的位点（如图 16-1 和图 16-2 中所示 RNA 聚合酶和活化子），事实也常常如此。但是有些互作蛋白质的结合位点在 DNA 上相距较远。在这种情况下，为了使蛋白质相互作用，结合位点之间的 DNA 就会形成环，让蛋白质结合位点彼此靠近（图 16-3）。

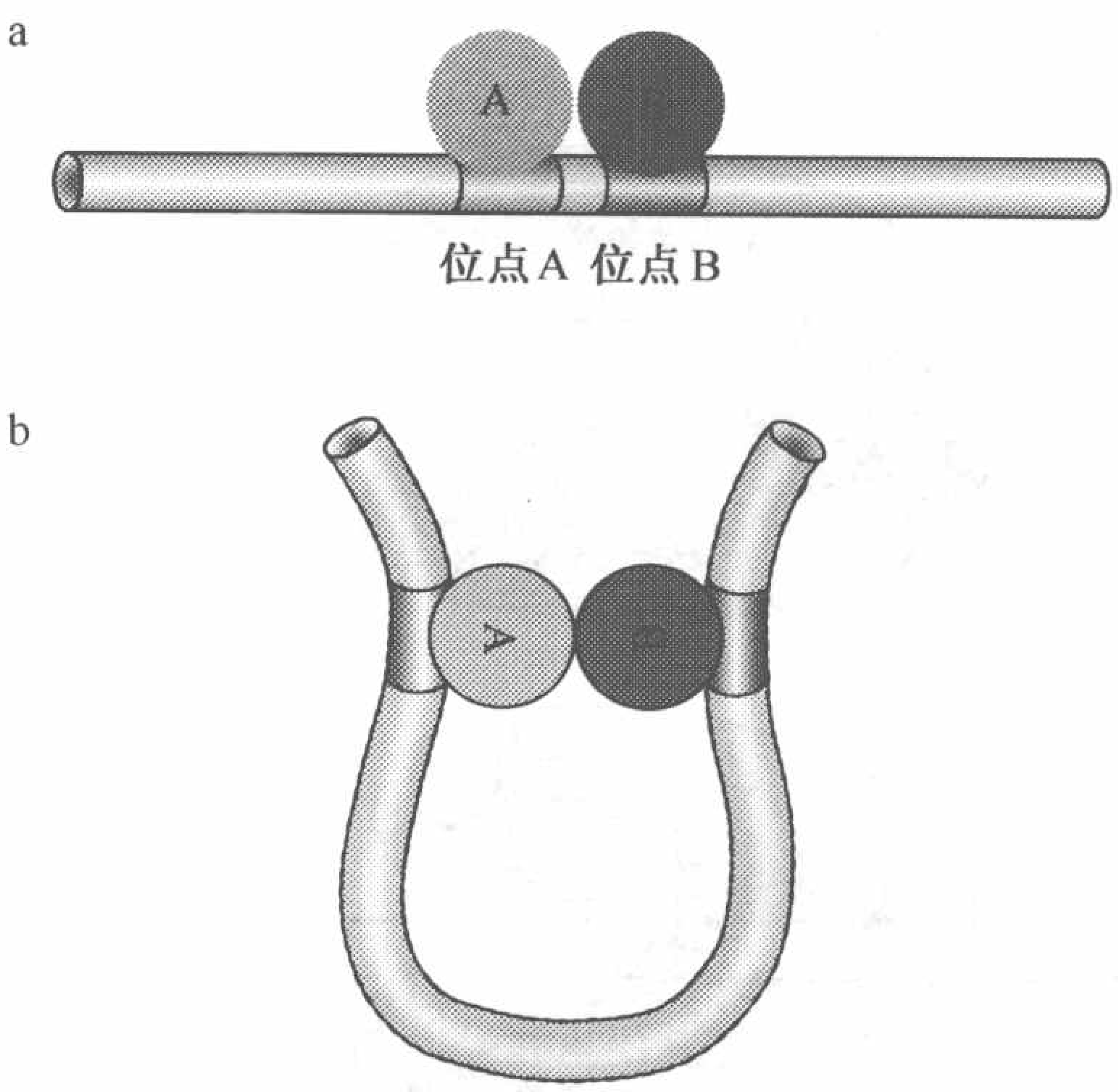


图 16-3 DNA 结合蛋白的相互作用。 a. 蛋白质协同结合相邻位点；b. 蛋白质协同结合分离的位点。

在细菌中我们遇到这种蛋白质的互作方式的例子。实际上，我们已经提及的 NtrC 活化子就是进行“远程”激活的：其结合位点一般位于启动子上游 150bp 处，在结合位点更远（1kb 或更多）的时候该活化子也能发挥作用。抑制子相互作用使 DNA 形成长达 3kb 的环的情况也不少见。在下一章（真核基因调控）的内容中，我们还将遇到更多、更生动的“远程激活”的例子。

使距离很远的 DNA 位点靠近以便 DNA 链环化的方式之一是在这些位点之间的

DNA 序列上再结合其他蛋白质。细菌中就有这样的例子，即在活化子结合位点和启动子之间另有蛋白质结合，它使 DNA 弯曲，协助活化子和聚合酶相互作用（图 16-4）。在其他过程中也有这种“构架型”的蛋白质协助蛋白质相互作用（如位点特异性重组，第 11 章）。

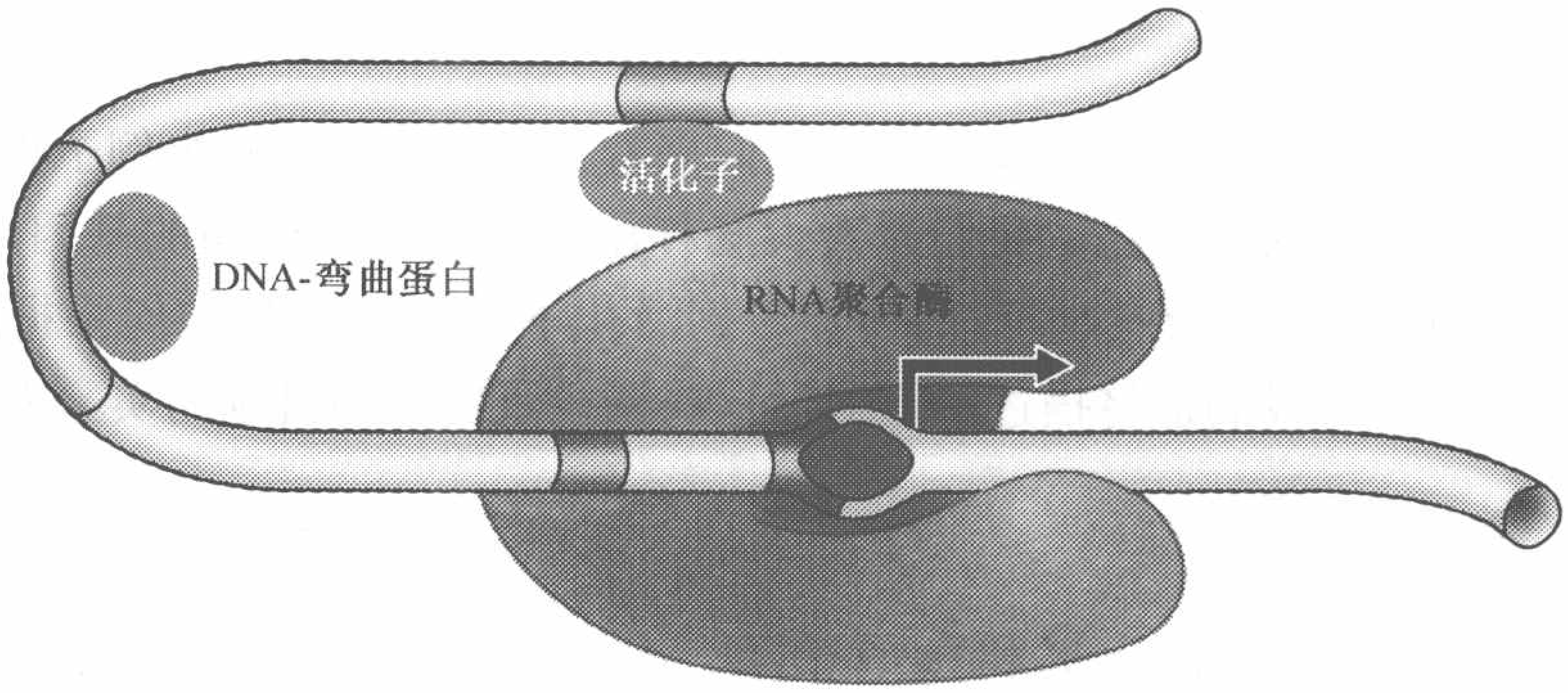


图 16-4 DNA-弯曲蛋白能够协助 DNA-结合蛋白间的相互作用。弯曲 DNA 的蛋白质结合到活化子结合位点和启动子之间的位点。这使得两个位点在空间上更接近从而协助 DNA-结合活化子和聚合酶的相互作用。

协同结合和变构在基因调控中有多种作用

我们已经指出简单的协同结合可以介导基因激活：活化子自发地与 DNA 和聚合酶相互作用从而募集 RNA 聚合酶到启动子上。另一方面，我们也描述了蛋白质变构作用介导激活转录的情形：活化子与结合到启动子上的聚合酶，通过诱导酶或启动子发生构象变化，刺激转录起始。在基因调控中，协同结合和变构还有其他作用。

例如，经常有成组的调控蛋白共同结合到 DNA 上，即两个或多个活化子和（或）抑制子彼此互作并与 DNA 相互作用，从而彼此协助结合到它们共同调控的基因附近。我们将看到，这种相互作用仅仅因为外界环境的小小变化就可以灵敏地使一个基因从完全不表达转而完全表达。活化子的协同结合还可以参与整合信号，也就是说，某些基因只有当存在多种信号（从而也就有多种调控蛋白）时才被激活。噬菌体 λ 中就有协同作用参与基因调控的非常典型的、目前人们知之甚详的例子。稍后我们将在本章介绍关于该例的内容和框 16-5 中详细讨论协同结合的基本机制及其影响。

变构不仅是一种基因激活的机制，它同时还常常是特异信号控制其调控蛋白的方式。迄今已知细菌调控蛋白通常采取两种构象：一种构象可以结合 DNA ，另一种则不能。信号分子结合调控蛋白并将它固定在某一特定构象，从而决定该调控蛋白是否能够激活转录。我们已经在第 5 章看到过一个这样的例子（图 5-25），当时我们还较为详细地考察了变构的基本机制。本章和下一章我们将看到几个信号通过变构控制其调控蛋白的例子。

抗终止作用及其延续：基因调控并不局限于转录起始调控

正如本章开始部分所述，基因调控主要发生在转录起始阶段，在真核细胞和细菌中

皆是如此。但是，这两类生物的基因调控决不局限在这一步骤。本章中我们会看到几个细菌中基因调控的例子，其中包括了转录延长、RNA 处理和 mRNA 翻译。

转录起始的调控：原核生物的实例

大致了解了转录调控的基本原理之后，我们来看几个实例，这几个例子将为我们展示基本原理是如何起作用的。首先，我们来看大肠杆菌中参与乳糖代谢的基因——*lac* 操纵子（*lac* operon）中的基因。在这里我们将会看到活化子和抑制子是如何根据两种不同的信号调控表达的。同时我们也将描述一下揭示这些调控蛋白作用机制的实验方法。

活化子和抑制子共同控制 *lac* 基因

3 个 *lac* 基因——*lacZ*、*lacY* 和 *lacA* 在 *E. coli* 基因组中相邻排列合称 *lac* 操纵子（图 16-5）。*lac* 启动子位于 *lacZ* 的 5′端，指导 3 个基因共同转录为一条 mRNA [因其包含多个基因故被称为多顺反子信使（polycistronic message）]。这一 mRNA 翻译产生 3 种蛋白质产物。*lacZ* 基因编码 β-半乳糖苷酶，该酶可将乳糖切割成半乳糖和葡萄糖，后两种糖均可用作细菌的能源。*lacY* 基因编码乳糖转移酶，该蛋白质插入细胞膜可以将乳糖转运到细胞内。*lacA* 基因编码硫代半乳糖苷转乙酰酶，它可以消除同时被 *lacY* 转入的硫代半乳糖苷对细胞造成的毒性。

这些基因只有在乳糖存在，同时葡萄糖（优先选用的能源）缺乏时才会高水平地表达。两种调控蛋白参与基因调控：一种是活化子称为 CAP，另一种是抑制子称为 Lac 抑制子（Lac repressor）。Lac 抑制子由 *lacI* 基因编码，这一基因位于其他 *lac* 基因的附近，但其转录（组成型表达）却是由其自身的启动子起始的。CAP 即为代谢活化子蛋白（catabolic activator protein），这一活化子也被称为 CRP（cAMP 受体蛋白，cAMP receptor protein）。编码 CAP 的基因位于细菌染色体的其他位点，不与 *lac* 基因连锁。CAP 和 Lac 抑制子都是 DNA 结合蛋白质，两者分别结合在 DNA 启动子上或其附近的特异位点上（图 16-5）。

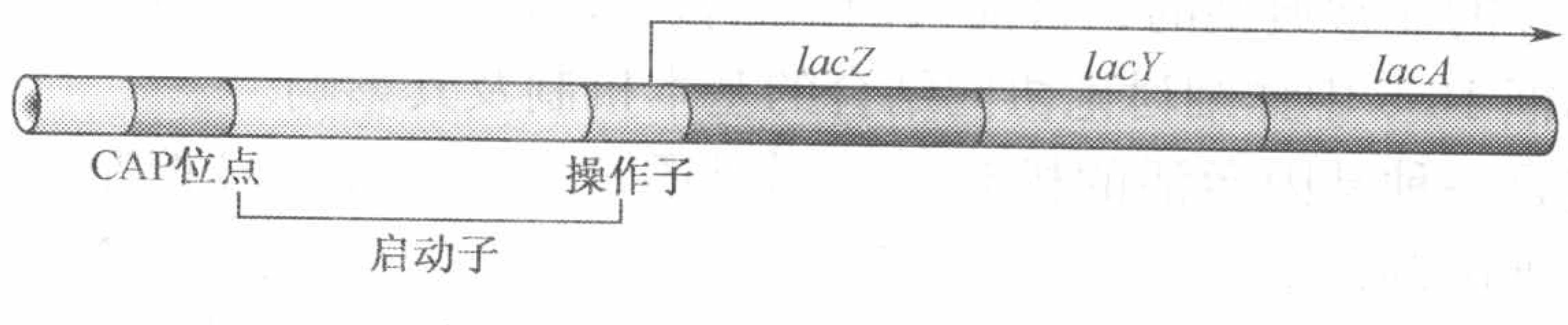


图 16-5 *lac* 操纵子。三个基因（*lacZ*、*lacY* 和 *lacZ*）以一条 mRNA 从启动子（如箭头所示）转录。CAP 位点和操作子各约为 20bp。操作子位于启动子 RNA 聚合酶结合区内，CAP 位点位于启动子上游（详见图 16-8 这些位点的相关排列及正文中关于这些位点的结合蛋白的描述）。附近还有两个额外的弱 *lac* 操作子（图 16-13），本图简化掉了，我们目前不必考虑它们。

每一个调控蛋白分别响应一种环境信号并将其传递给 *lac* 基因。CAP 介导葡萄糖信号，而 Lac 抑制子介导乳糖信号。这一调控系统以如下方式作用：Lac 抑制子只在乳糖缺

乏的时候才结合 DNA 从而抑制转录，在乳糖存在时，抑制子没有活性，基因是去抑制（表达）的。CAP 只在没有葡萄糖的时候才结合 DNA 从而激活 *lac* 基因。这样两种调控蛋白联合作用保证基因在乳糖存在，同时缺少葡萄糖的环境中以显著的水平表达（图 16-6）。

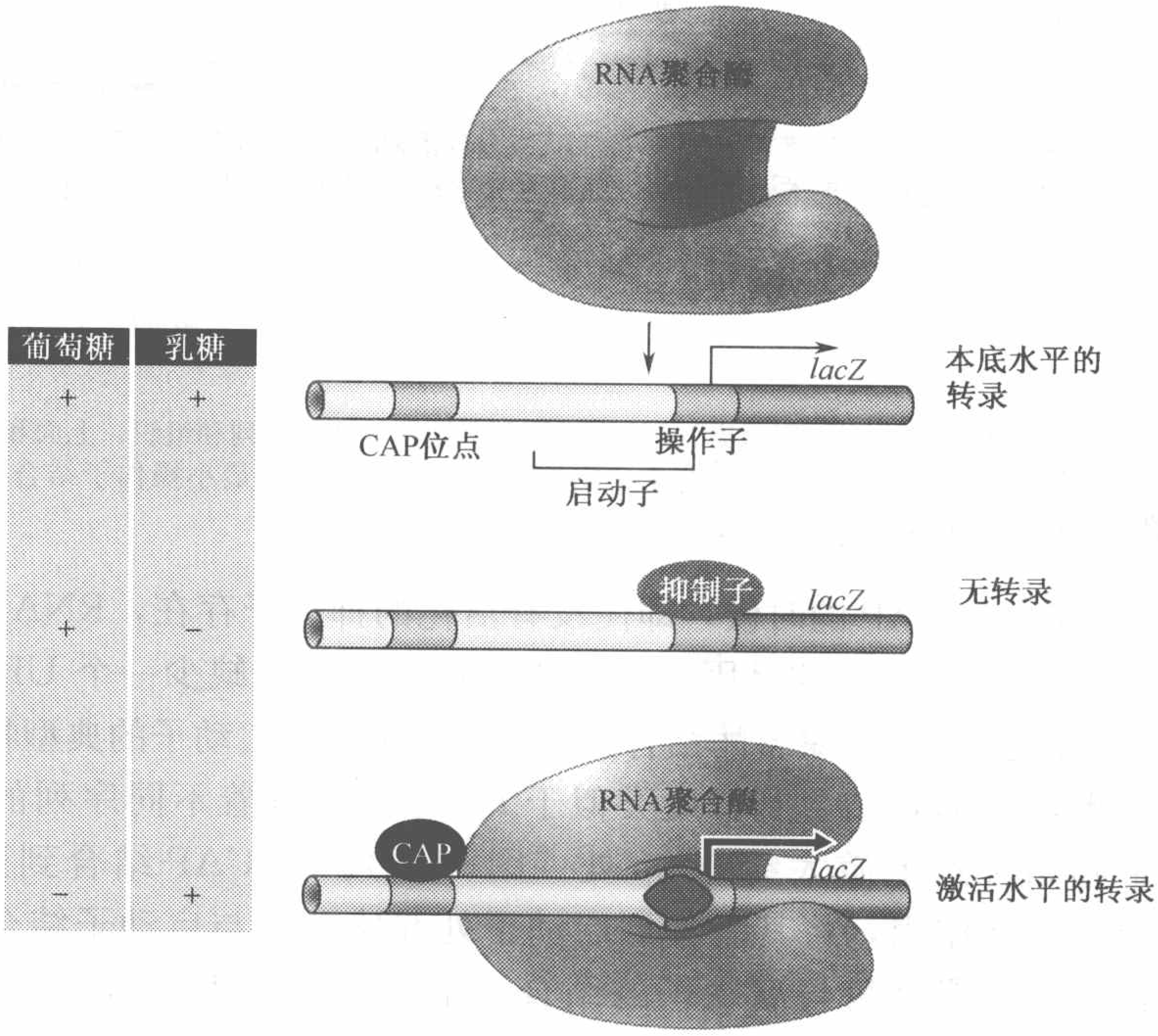


图 16-6 *lac* 基因的表达。乳糖和葡萄糖的有无控制着 *lac* 基因表达的水平。高水平表达要有乳糖存在（也就是没有有功能的 Lac 抑制子）同时缺少偏好的能源葡萄糖（因此也就有活化子 CAP）。当 Lac 抑制子结合到操作子上，不管活化的 CAP 存不存在，它都排斥聚合酶。图示的 CAP 和 Lac 抑制子都是单亚基。但实际上，CAP 以二聚体结合 DNA，Lac 抑制子以四聚体结合（图 16-13）。CAP 募集聚合酶到 *lac* 启动子，在这里，它自动的异构化为开放复合体。（最底行所示的状态）

CAP 和 Lac 抑制子对 RNA 聚合酶结合 *lac* 启动子的作用是相反的

我们已经看到，Lac 抑制子结合的位点被称为 *lac* 操作子（*lac* operator）。这一双重对称的 21bp 序列被 Lac 抑制子的两个亚基识别，每一个结合一个“半位点”（图 16-7）。本章后面的“CAP 和 Lac 抑制子结合 DNA 时使用相同的结构模体（motif）”部分我们会详细地描述它们是如何结合的。当抑制子结合到操作子上时，它是怎样抑制转录的呢？

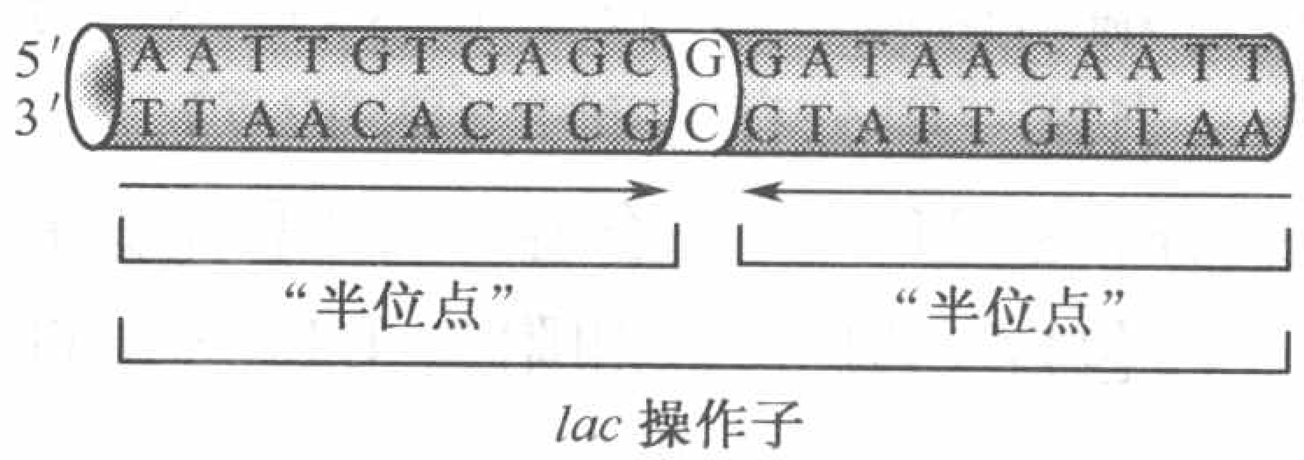


图 16-7 *lac* 操作子对称的半位点。

lac 操作子与启动子有重叠区，所以当抑制子结合到操作子上时，就阻止了 RNA 聚合酶在启动子上的结合和 RNA 合成的起始（图 16-8）。用框 16-1 中所述的 DNA 足迹法和凝胶迁移率分析可以确认蛋白质结合位点并对它们的位置作图。

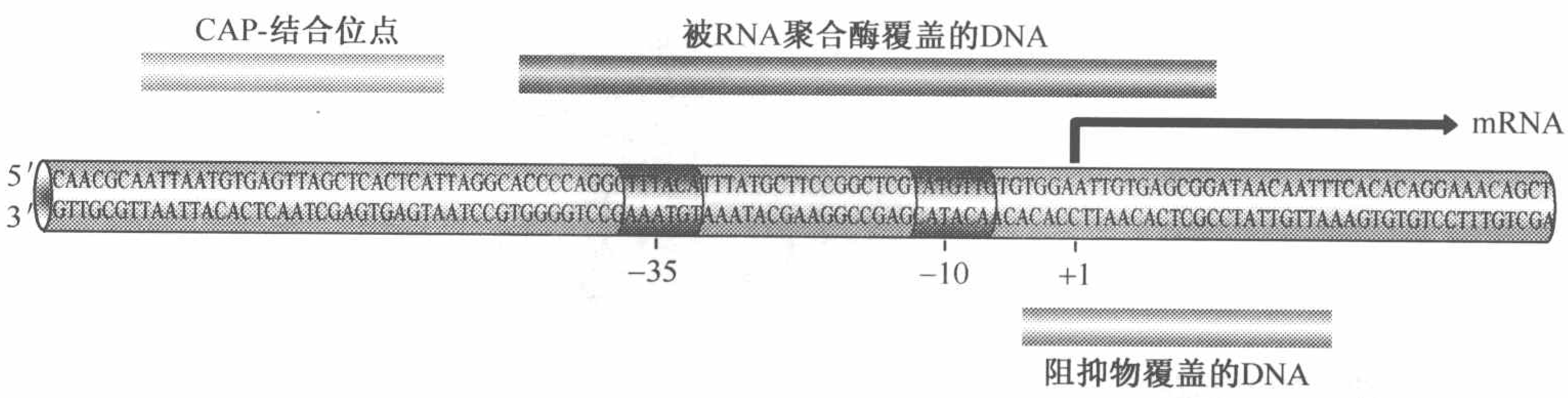


图 16-8　*lac* 操纵子的控制区域。图中所示为 *lac* 操纵子控制区域的核苷酸序列和组织。上面和下面的横杠表示 RNA 聚合酶和调控蛋白覆盖的 DNA。注意 Lac 阻抑物覆盖的区域大于最小操作子结合位点序列，而 RNA 聚合酶覆盖的区域大于启动子的组成序列。

我们看到，在没有 CAP 的时候，即使没有活化的抑制子存在，RNA 聚合酶与 *lac* 启动子也只有很微弱的结合。这是因为 *lac* 启动子的 −35 区缺少一个 UP 成分（第 12 章和图 16-8）而不适合 RNA 聚合酶结合，这是活化子控制启动子的典型特征。

CAP 以二聚体的形式结合到长度类似于 *lac* 操作子有着不同序列的位点上。该位点位于转录起始位点上游大约 60bp 处（图 16-8）。当 CAP 结合到这一位点时，它就会与聚合酶相互作用，将其募集到启动子处并协助其与启动子结合（图16-6），这一协同结合稳定了聚合酶与启动子的结合。下面我们详细地看一下 CAP 介导的转录激活。

CAP 具有独立的激活表面和 DNA 结合表面

各种实验都支持这一观点，即 CAP 通过募集 RNA 聚合酶激活 *lac* 基因。分离到的 CAP 的突变体结合 DNA 但并不激活转录。这些所谓正控制（positive control）突变体的存在表明，要激活转录，活化子不单要结合在启动子附近的 DNA 上，转录激活也不是由活化子改变局部 DNA 的结构引起的。正控制突变体的氨基酸替换确认了 CAP 接触聚合酶的区域，称之为激活区（activating region）。

当 CAP 激活 *lac* 基因时，其激活区接触的是 RNA 聚合酶的哪一部分呢？聚合酶的突变形态揭示了这一位点，这种突变体聚合酶可以正常地转录多种基因，但是不能在 *lac* 基因处被 CAP 激活。这些突变体在 RNA 聚合酶的 α 亚基（α subunit）的 C 端结构域（C-terminal domain）发生了氨基酸替换。我们在第 12 章已知，这一结构域由一个柔性的衔接体连接到 α 的 N 端结构域（NTD）。α NTD 深埋于酶内部，而 α CTD 却延伸出来与启动子组分结合（当这一组分存在时）（图 12-7）。

lac 启动子没有 UP 成分，α CTD 就结合在 CAP 和附近的 DNA 上（图 16-9）。包括了 CAP、α CTD 以及一段包含 CAP 位点和附近 UP 成分的 DNA 寡核苷酸的复合体的晶体结构证明了本结构（图 16-10）。在框 16-1 中，我们将描述一个实验，该实验显示 *lac* 启动子的激活只需要聚合酶的募集。

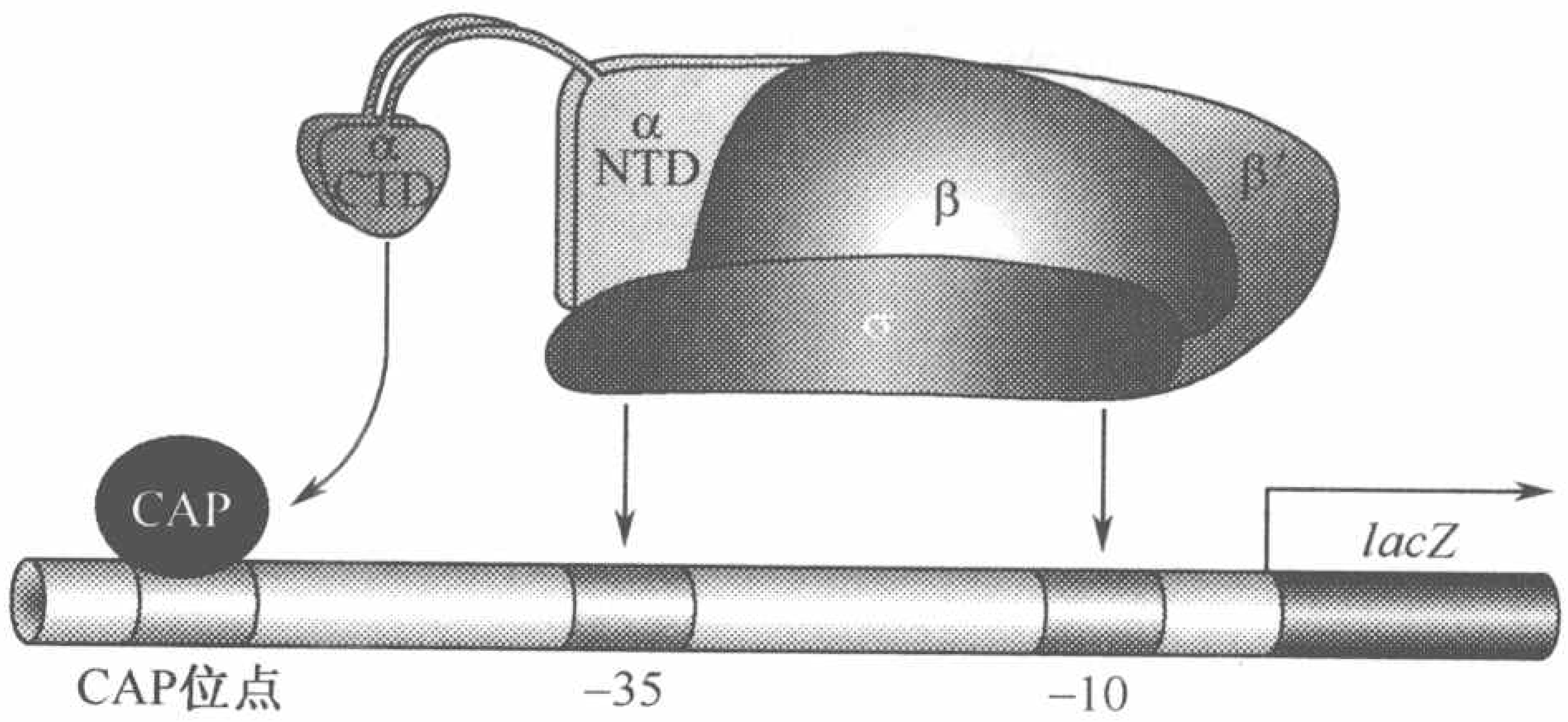


图 16-9 *lac* 启动子被 CAP 激活。RNA 聚合酶在 CAP 的协助下结合到 *lac* 启动子。CAP 由 α 亚基的 CTD 识别。当与 CAP 相互作用时，α CTD 也在靠近 CAP 位点处与 DNA 接触。在第 12 章，我们在说明活化子和其聚合酶上的目标位点或聚合酶与启动子区域的接触时，也使用了 RNA 聚合酶作为代表。

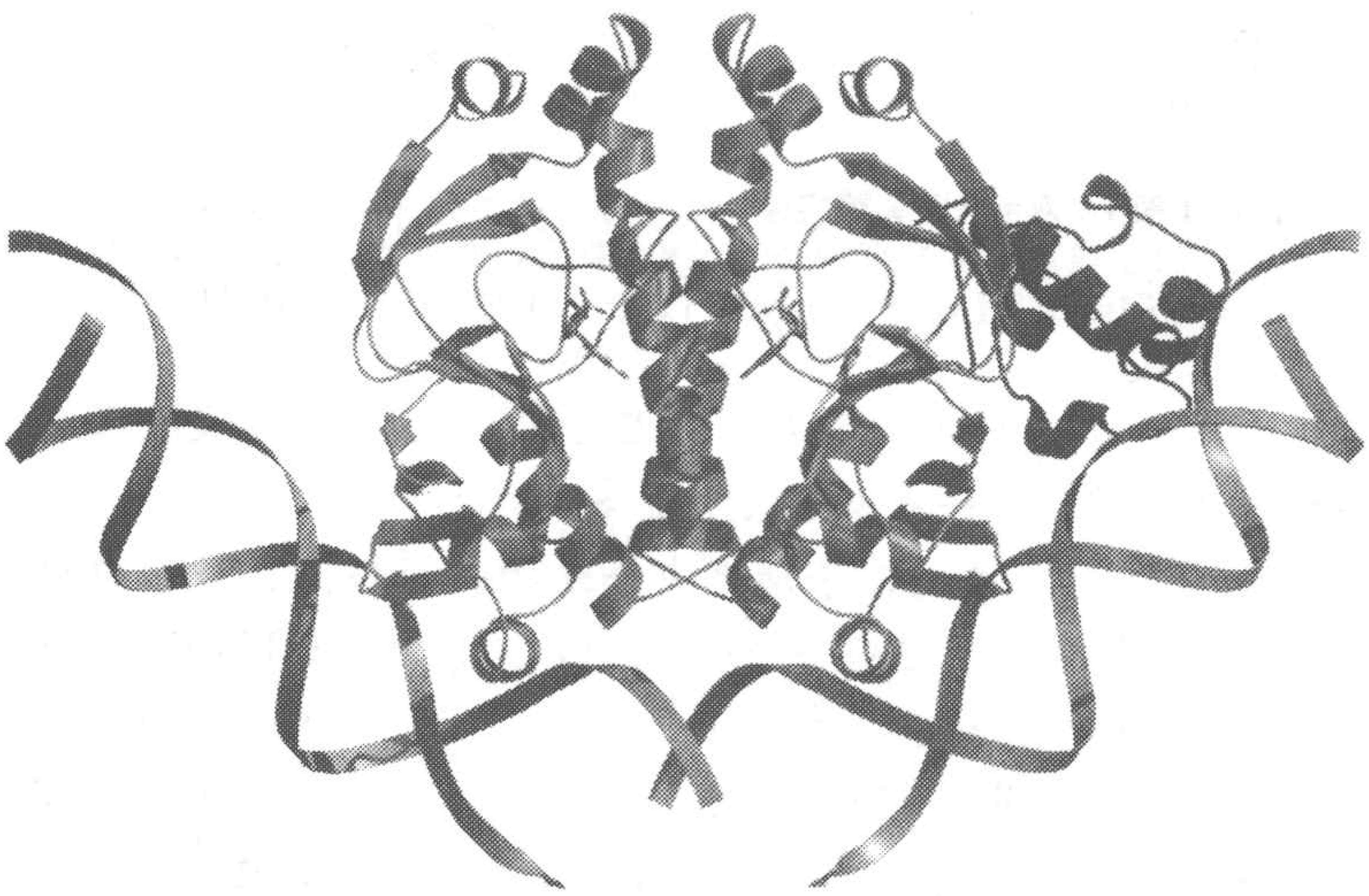
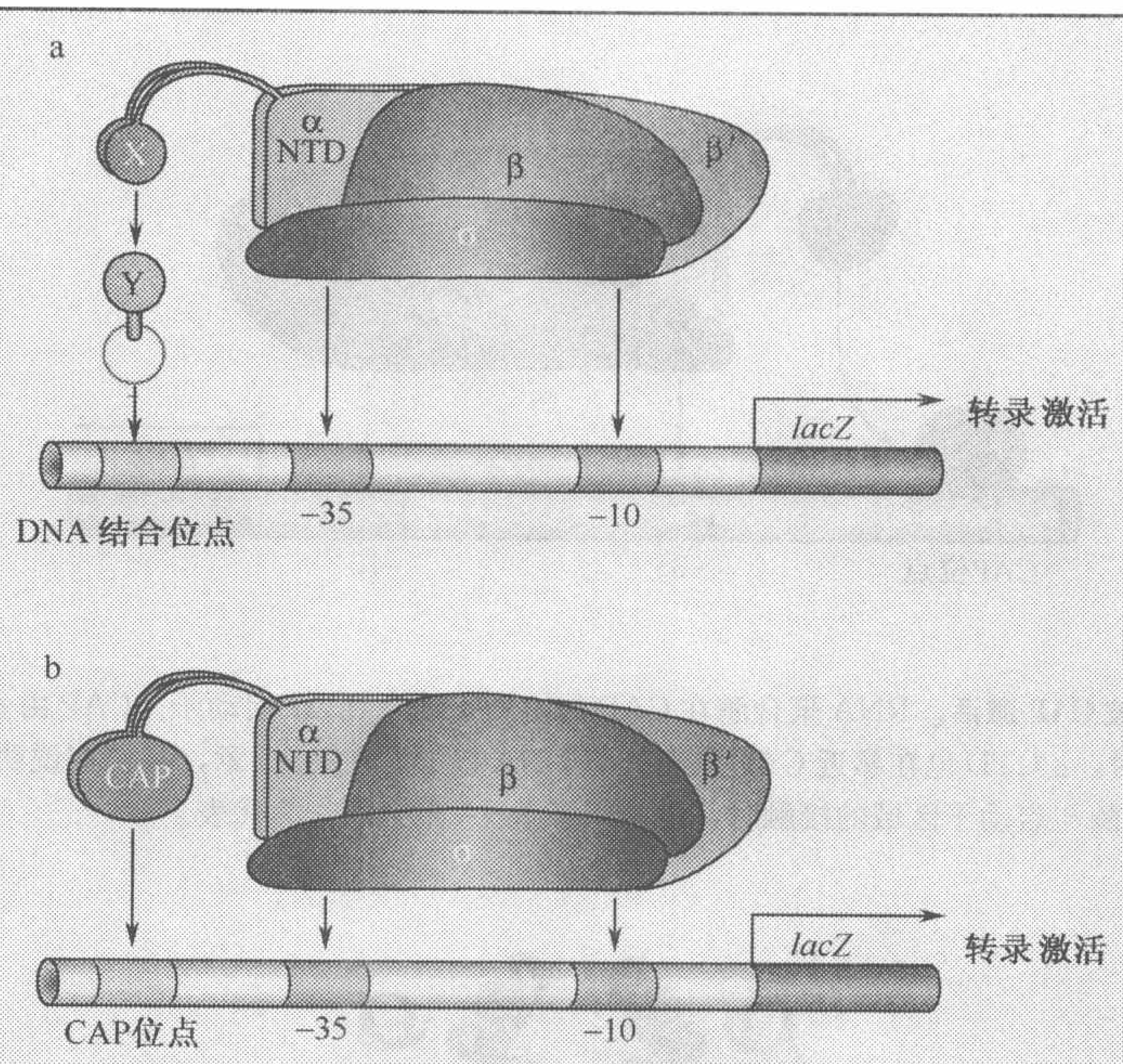


图 16-10 CAP-α CTD-DNA 复合物的结构。所示的 CAP 以二聚体结合到其位点上，如我们在图 15-8 中所见的。而且，在本例中，RNA 聚合酶的 α CTD 结合到一段 DNA 上，与 CAP 相互作用。每个蛋白质相互作用的位点包含用遗传学方法确认的残基。在本图中，CAP 以绿色表示，α CTD 以紫色表示。图中所示每一个 CAP 还结合了一分子的 ATP。(Benoff B. et al. 2002. Science 297：1562.）图像用 MolScript 和 Raster 3D 软件制作。

在了解了 CAP 怎样激活 *lac* 操纵子的转录以及 Lac 抑制子怎样拮抗这一作用之后，我们再进一步看一下这些调控蛋白是如何识别其 DNA 结合位点的。

框 16-1 活化子旁路实验

如果活化子只具有募集聚合酶到基因上的功能，那么将聚合酶带到基因上的方法也可以同样起作用。对 *lac* 基因来说确是如此，如以下实验所示（框 16-1 图 1）。



框 16-1 图 1　双活化子旁路实验（a）与蛋白质 Y 相互作用的 X 蛋白质替代 α CTD。蛋白质 Y 与 DNA 结合结构域融合，被这个结构域识别的位点处于临近 Lac 基因的位置。(b) α CTD 被 CAP 的 DNA 结合区域所替代。

在一个实验中，用另一蛋白质-蛋白质相互作用取代 CAP 和聚合酶的相互作用。用两个已知的蛋白质相互作用，一个与 DNA 结合结构域联合，另一个取代聚合酶 α 亚基的 C 端结构域（α CTD)。只要启动子附近有合适的结合位点，被修饰过的聚合酶就能被拼凑起来的“活化子”激活。在另一个实验中，聚合酶 α CTD 被 DNA 结合结构域取代（如 CAP 的 DNA 结合结构域)。在无任何活化子时，只要附近有合适的 DNA 结合位点，这一修饰过的聚合酶就能有效地起始从 *lac* 开始的转录。第三个实验更简单：聚合酶能在体外无任何活化子时高水平的转录 *lac* 基因，但要求聚合酶是高浓度的。这样我们看到，人工募集或提供高浓度的聚合酶都可以使 *lac* 基因以激活水平表达。这些实验与活化子协助聚合酶结合启动子是一致的。关于为何仅仅提高蛋白质（如 RNA 聚合酶）的浓度就可以协助它结合到 DNA 上的位点（在这里是启动子）的解释见框 16-5。其中讨论的结果出人意料。活化子通过诱导聚合酶发生特异的变构来激活转录。

CAP 和 Lac 抑制子以相同的结构模体结合 DNA

细菌的许多活化子和抑制子包括 CAP 和 Lac 抑制子结合 DNA 的结构基础已经用 X 射线晶体学确定。细菌调控蛋白的 DNA 识别机制大多是相似的，仅在细节上有所不同，正如我们下面将要描述的。

一般情况下，蛋白质以同源二聚体的形式结合到反向重复（或近似重复）序列的位点上。每一个单体结合到一个“半位点”，二聚体的对称轴位于结合位点的对称轴之上（就如我们在 Lac 抑制子中所见到的，图 16-7)。蛋白质靠一种保守的二级结构螺旋-转

折-螺旋（helix-turn-helix，图 16-11）识别特异的 DNA 序列。这一结构域由两个 α 螺旋组成，其中一个是识别螺旋（recognition helix）进入 DNA 的主凹槽（大沟）。在第 5 章我们就讨论过，α 螺旋的大小正好适合进入 DNA 主凹槽，利于其外表面的氨基酸残基与碱基对边缘的化学基团相互作用。请回忆第 6 章我们看到的每一种碱基有着特有的氢受体和氢供体结构模式（图 6-10）。这样，蛋白质不必解 DNA 双螺旋链也可以区别不同的 DNA 序列（图 16-11）。

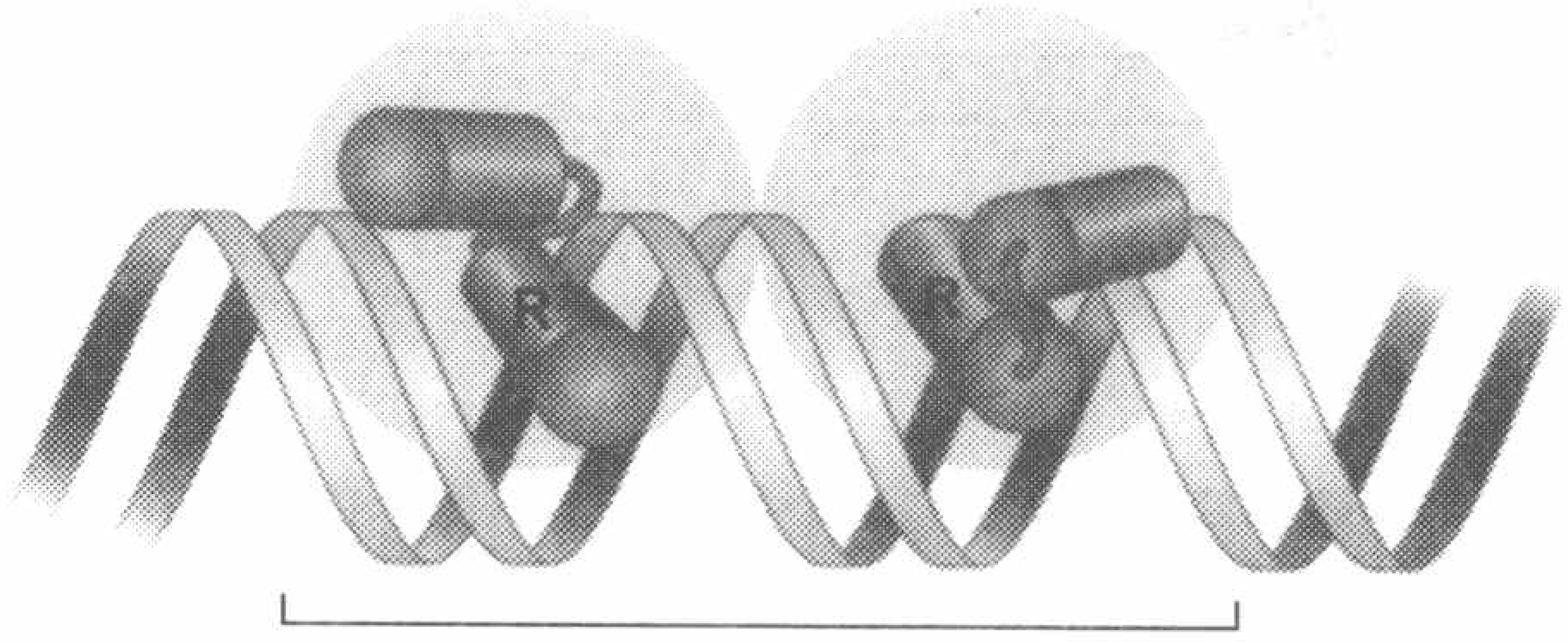


图 16-11 具有螺旋-转折-螺旋结构域的蛋白质结合 DNA。蛋白质就像实际情况中那样以二聚体结合，两个亚基用带有阴影的球表示。标示出了每个单体的螺旋-转折-螺旋模体。“识别螺旋”标以 R。

识别螺旋伸出的氨基酸侧链和碱基边缘的联系可以由直接或间接（由水分子搭桥）的 H 键或者范德华力介导。我们在第 3 章讨论过这些键的性质，在第 5 章和第 6 章讨论了它们在 DNA 识别中的作用。图 16-12 中所示的是某种识别螺旋和其 DNA 结合位点的相互作用的例子。

螺旋-转折-螺旋结构域的第二个螺旋坐落的位置横跨 DNA 大沟与 DNA 主链相联系，以保证识别螺旋出现在正确的位置，同时也增加整个蛋白质-DNA 相互作用的结合能。

这里描述的情况不仅与 CAP（图 5-18 和图 16-10）和 Lac 抑制子的情况本质上符合，而且细菌中许多调控蛋白也是如此，包括我们将在稍后遇到的 λ 噬菌体抑制子和 Cro 蛋白。它们在细节上会有些不同，如下面的例子所示：

- Lac 抑制子以四聚体而不是二聚体结合。但是，每一个操作子只与这 4 个亚基中的两个接触。所以，这种寡聚方式的不同

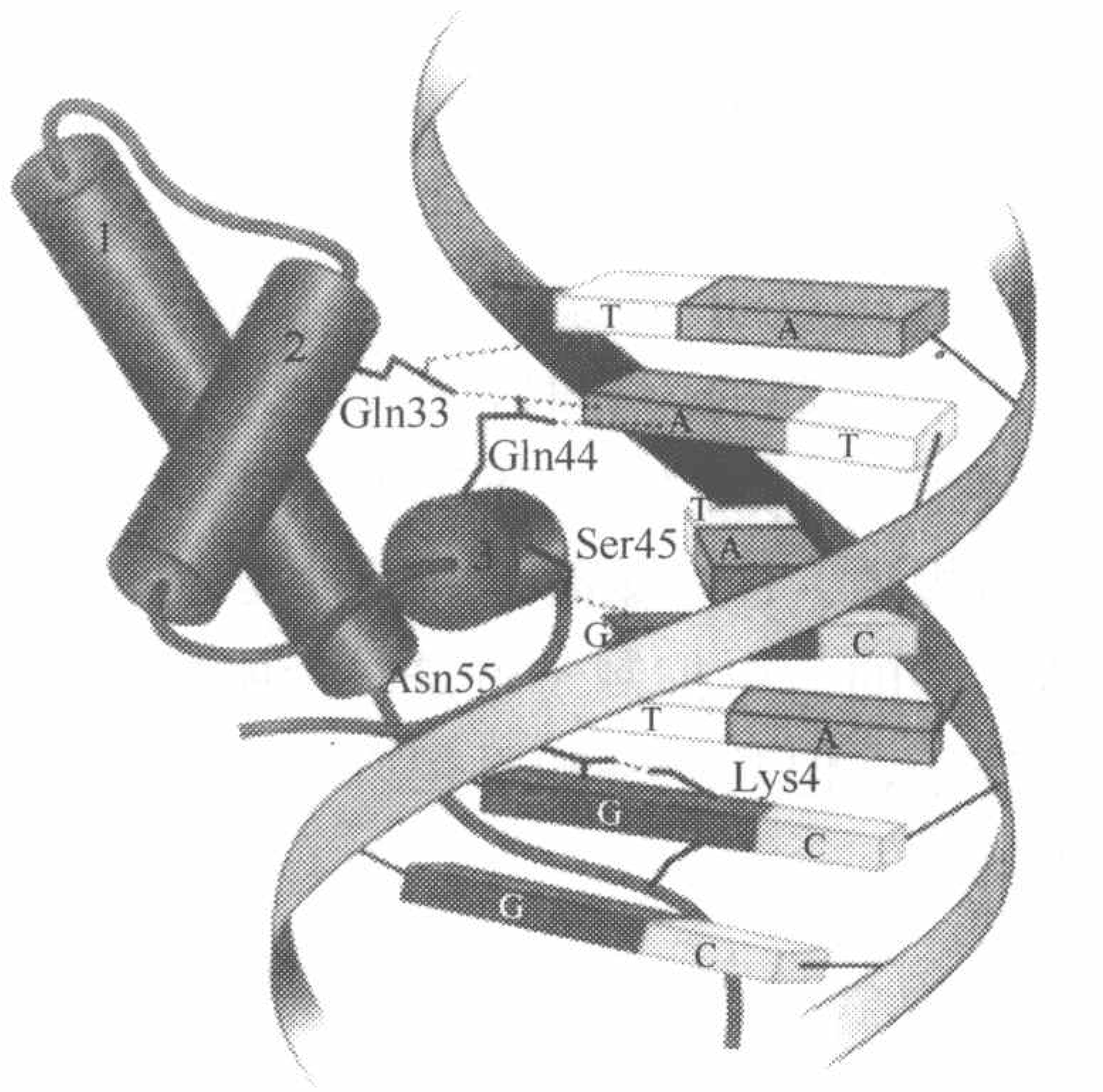


图 16-12 操作子大沟处 λ 抑制子和碱基对之间的氢键。抑制子-操作子复合物图，显示了侧链上氨基酸和半位点中的碱基之间的氢键（用虚线表示），只表示了相关的氨基酸和侧链。在识别螺旋中除了 Gln44 和 Ser45，紧接识别螺旋的环中的 Asn55 也与特异的碱基接触。而且（这种情况比较少见，见稍后的正文）在蛋白质的 N 端中的 Lys4 也在大沟中与 DNA 螺旋的背面接触。Gln33 与骨架接触。（来源：Redrawn from Jordan，S. and Pabo，C. Science 242：896，Fig. 3B.）

并不会改变 DNA 识别的机制。四聚体中的另外两个单体可以再结合两个操作子，这两个操作子一个位于初级操作子下游 400bp 处，一个位于初级操作子上游 90bp 处。这时，位于中间的 DNA 就处于环化状态利于反应的进行（图 16-13）。

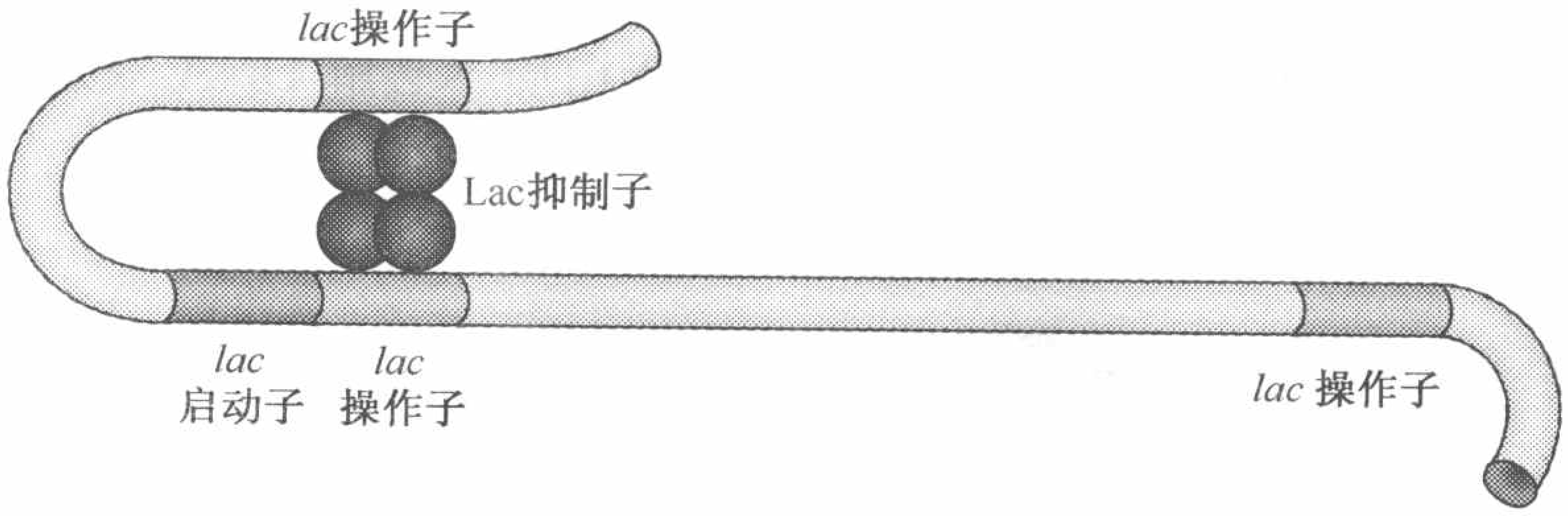


图 16-13 Lac 抑制子以四聚体结合到两个操作子上。图示的环是在 Lac 抑制子结合的初级操作子和上游的辅助操作子之间，也可与下游操作子选择性的形成类似的环。初级操作子（与启动子相对的那个）是在讨论 *lac* 基因表达时所指的操作子。在本图中，每一个抑制子都以两个球表示，而不是一个椭圆。（如前面图中所用的）

- 有时，在螺旋-转折-螺旋结构域之外的其他蛋白质区域也会与 DNA 相互作用。例如，λ 抑制子就以其 N 端臂与 DNA 接触。这些蛋白质区域环绕 DNA 在螺旋的背面与小沟相互作用（图 16-12）。
- 许多情况下，蛋白质的结合不会改变 DNA 的结构。但是，有时，我们也会在蛋白质-DNA复合物中看到各种 DNA 弯曲。例如，CAP 半包围着 DNA 并诱导 DNA 产生巨大的弯曲。这是由螺旋-转折-螺旋结构域外的蛋白质区域与启动子之外的序列相互作用引起的。有时，结合也会导致操作子 DNA 扭曲。

原核生物的抑制子并不都使用螺旋-转折-螺旋的方式与 DNA 结合。其他的结合方式也见诸报道。P22 噬菌体［一种与 λ 噬菌体相近，但感染沙门氏菌（*Salmonella*）的噬菌体］中的 Arc 抑制子就是一个明显的例子。Arc 抑制子以二聚体结合一个反向重复序列操作子，但其亚基却不是以 α 螺旋结构，而是以两个反平行 β 链插入大沟识别结合位点的。

Lac 抑制子和 CAP 的活性由其信号变构控制

当乳糖进入细胞后，它就被转变为别乳糖，正是别乳糖（而不是乳糖自身）控制 Lac 抑制子。与此相矛盾的是，乳糖向别乳糖的转变是由 β-半乳糖苷酶催化，该酶自身也是由 *lac* 基因中的一个编码的，这怎么可能呢？

答案就是 *lac* 基因表达的渗漏性：在它们被阻抑时，也会偶尔被转录，这时因为 RNA 聚合酶偶尔会取代 Lac 抑制子成功地结合到启动子上。这种渗漏保证了细胞中在没有乳糖的时候也会有低水平的 β-半乳糖苷酶，从而可以催化乳糖转化为别乳糖。

别乳糖结合 Lac 抑制子引起其蛋白质形状（构象）的改变。在没有别乳糖时，抑制子以可以结合 DNA（从而关闭 *lac* 基因）的形式出现，而一旦别乳糖改变了抑制子的形状，蛋白质就不可以再结合 DNA，*lac* 基因就不再被抑制。在第 5 章中我们描述过

Lac 抑制子（图 5-25）这一变构的结构基础。需要强调的重要一点是，别乳糖结合 Lac 抑制子的非 DNA 结合结构域。

CAP 活性的调节方式与此相似。葡萄糖降低细胞内的一种小分子——cAMP 的浓度，这一分子是 CAP 的变构因子：只有当 CAP 与 cAMP 形成复合物时，CAP 蛋白才会形成结合 DNA 的构象。因此，只有当葡萄糖水平低（而 cAMP 水平高）时，CAP 才会结合 DNA，激活 *lac* 基因。CAP 蛋白结合 cAMP 的部分和蛋白质结合 DNA 的部分是分开的。

E. coli 的 *lac* 操纵子是早期法国生物学家 Francois Jacob 和 Jacques Monod 用来阐述基因调控的两个系统之一。在框 16-2 中，我们就这一方面的早期研究和这些观点何以会产生如此大的影响做了说明。

框 16-2　Jacob、Monod 及其基因调控的观点

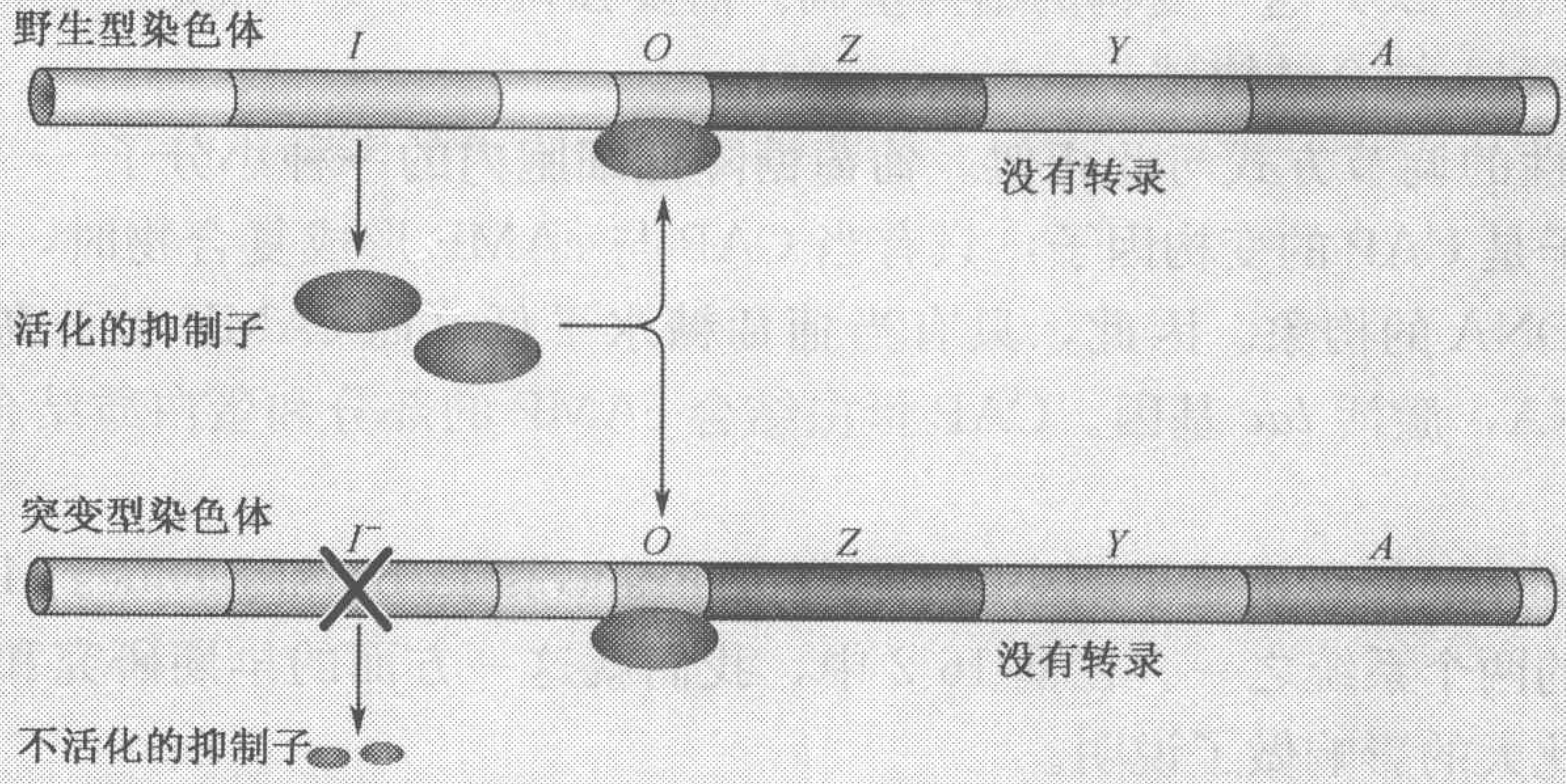
某一基因的表达可以由另一基因的产物控制，即存在调控基因，其产物唯一的功能就是调控其他基因的表达。这一观点是分子生物学早期的远见卓识之一。它是由一群在巴黎工作的科学家特别是 Francois Jacob 和 Jacques Monod 于 20 世纪 50 年代至 60 年代早期提出的。他们当时试图解释两种看起来不相关的现象：生长于乳糖的 *E. coli* 中 β-半乳糖苷酶的出现和细菌病毒（噬菌体）λ 在感染时的表现。他们工作的顶点是 1961 年其操纵子模型的发表以及 1965 年与他们的同事，Andre Lwoff 共同分享了诺贝尔生理学或医学奖。

我们今天难以理解他们取得的成就的伟大之处，因为我们对他们的观点如此熟悉，而且其模型的测试方法又是如此直接。为了理解这一点，就要考虑到他们开始其经典实验的时候已知些什么：只有在培养基中供给乳糖时，*E. coli* 中才会出现 β-半乳糖苷酶活性。当时尚不清楚该酶的出现需要打开一个基因的表达。实际上，早期的解释是细胞包含一个通用酶，这个酶根据环境的需要会具有各种属性。而当乳糖出现时，通用酶就会利用乳糖作为模板，获得合适的形态，代谢乳糖。

Jacob、Monod 和他们的同事用遗传学方法解析了这个问题。我们不再详细讨论他们的实验，而是做一个简单概述，以一窥其独创性。

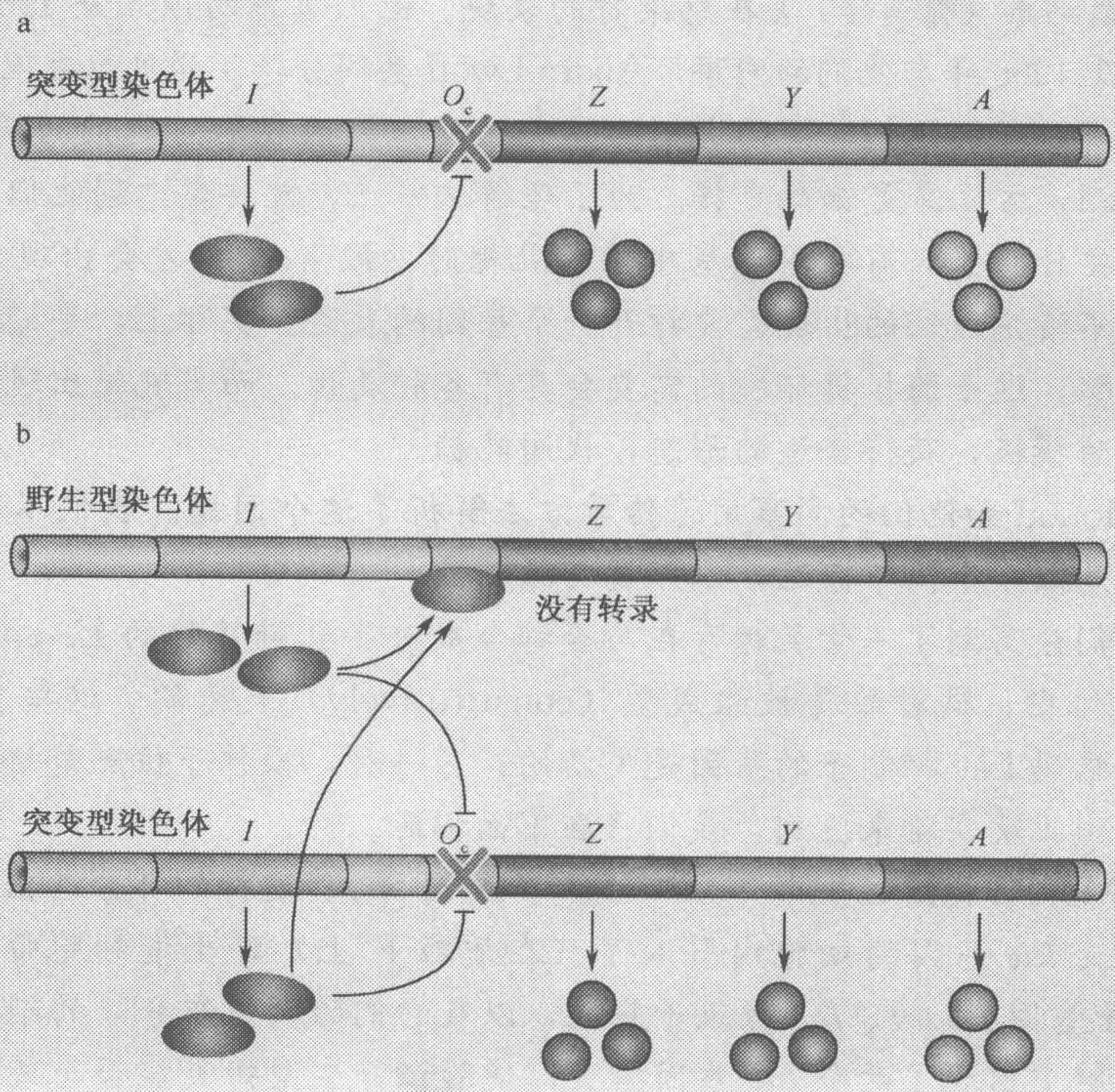
首先，他们分离到了不管乳糖存在与否都合成 β-半乳糖苷酶的 *E. coli* 突变体，即在这样的突变体中，这种酶是被组成型（constitutively）表达的。这些突变体分为两种：一种，其编码 Lac 抑制子的基因是失活的；另一种，操作子位点是缺陷型的。这两种突变体可以由正反实验来区别，我们将在下面描述。

Jacob 和 Monod 构建了一个部分二倍体（第 21 章）。这样的部分二倍体是向自身染色体上带有突变 *lac* 基因的细胞内引入了（在质粒 F′ 上）野生型细胞带有的 *lac* 基因（即 Lac 抑制子基因、*lacI*、*lac* 操纵子基因，以及它们的调控组分）的染色体片段。这种转化导致细胞中出现了两个 *lac* 基因拷贝，使验证野生型拷贝能否与某一特定的突变拷贝互补成为可能。当染色体基因由于 *lac I* 基因（编码抑制子）的突变而组成型表达时，质粒上的野生型拷贝可以恢复抑制（和可诱导性）。例如，β-半乳糖苷酶恢复只在乳糖存在时合成（框 16-2 图 1）。这是因为由质粒上的 *lac I* 基因合成的抑制子可以扩散到染色体上，即它可以反式作用。



框 16-2 图 1 部分二倍体细胞显示有功能的抑制子顺式工作。在没有乳糖时，Lac 基因不表达，所以在这些细胞中没有明显水平的 β-半乳糖苷酶的表达。

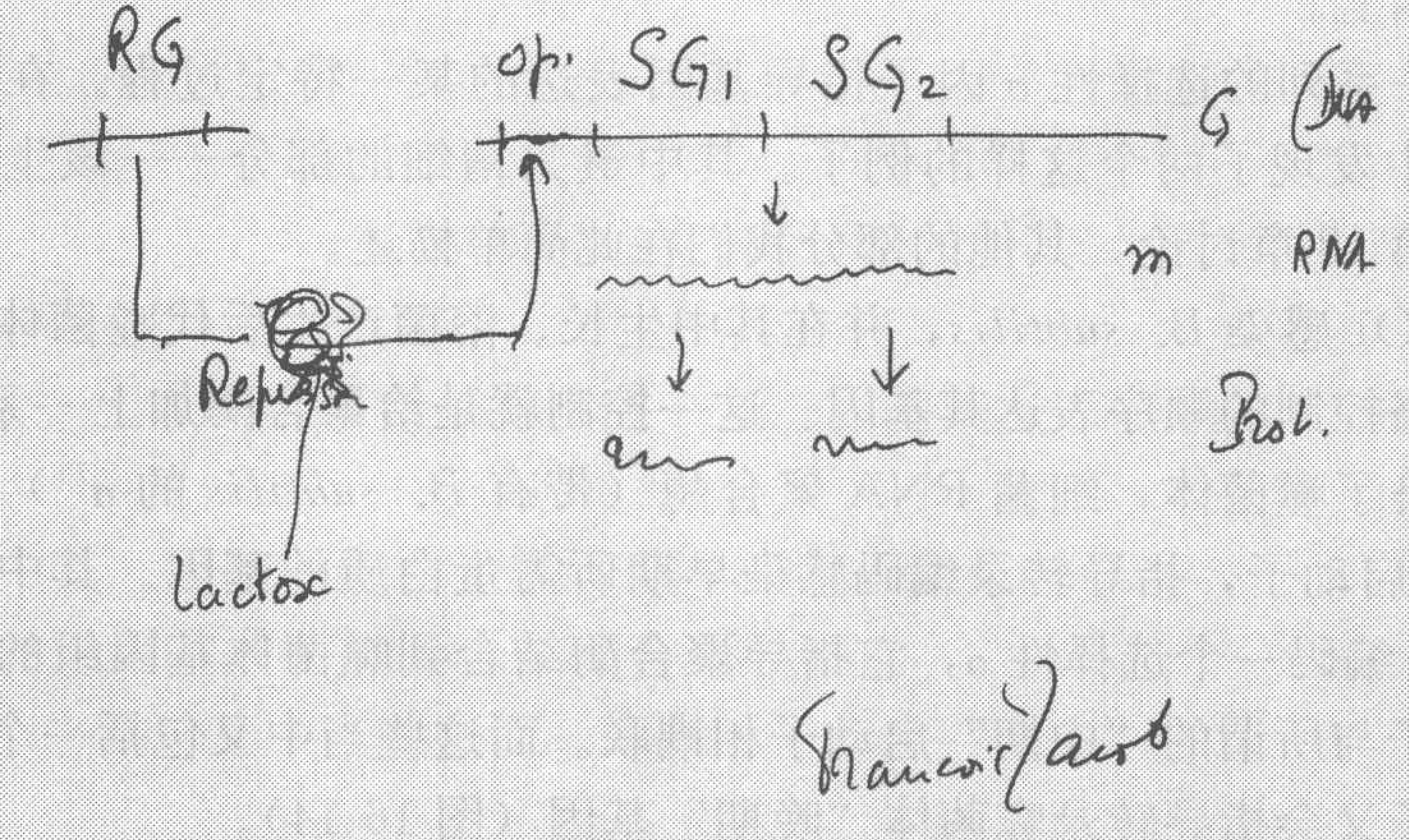
当造成染色体基因组成型表达的突变是在 *lac* 操作子上时，就不能被野生型基因反式互补了（框 16-2 图 2）。操作子只能顺式作用，即它只作用于与它在同一 DNA 分子上连锁的基因。



框 16-2 图 2 部分二倍体细胞显示操作子只以反式方式表达。（a）含有突变型操作子的单倍体细胞（O_c）。（b）部分二倍体细胞含有正常的操作子（O）和突变型操作子（O_c）。Lac 基因（Z、Y 和 A）附着在突变型操作子上持续的组成型表达，即使在同一细胞中另一染色体上有野生型操纵子也是如此。所以，操纵子只以反式工作。

这些实验连同其他实验的结果使 Jacob 和 Monod 提出基因在基因开始处称为启动子的特异位点开始表达，这种表达受抑制子的调控，抑制子通过位于启动子旁的操作子位点作用。

但是这些关于 *lac* 系统的实验不是单独进行的；Jacob 和 Monod 还同时用 λ 噬菌体（本章我们还将详细讨论这一系统）做了类似的实验。λ 噬菌体可通过两种生活周期增殖。选择哪种增殖方式依赖于表达了哪种相关的基因。这两位法国科学家发现，和 *lac* 基因一样，在 λ 噬菌体系统中也可以分离到控制基因表达的缺陷型突变体。明确了这些突变也是通过顺式操作子位点作用的反式作用抑制子。这两个调控系统（尽管是不同生物）的相似性使 Jacob 和 Monod 相信他们发现了基因调控的基本机制，而他们的模型适用于整个自然界。正如我们看到的，尽管他们的描述尚不完整[最值得注意的是，他们的方案中没有包括活化子（如 CAP）]，但他们提出的关于反式调控蛋白识别顺式调控位点的模型主导了后来的基因调控的研究。



框 16-2 图 3　这幅图显示 Lac 操作子及其调控，是由 Francois Jacob 在 2002 年手绘的。（由 Jan Witkowski 赠送）

组合控制：CAP 也控制其他基因

lac 基因是一个信号整合（signal integration）的很好例子：它们的表达受两种信号的控制，每种信号通过一种调控蛋白——Lac 抑制子和 CAP 将信号传递给基因。

下面我们来看 *E. coli* 中的另一组基因——*gal* 基因。这些基因编码参与半乳糖代谢的酶。与 *lac* 基因一样，*gal* 基因也是只在其作为其底物的糖（这种情况下是半乳糖）存在，而没有其优先选用的能源葡萄糖时才表达。同样，与 *lac* 类似，两种信号也是通过两种调控蛋白——活化子和抑制子传递给基因的。抑制子由 *galR* 编码，介导诱导物半乳糖的作用；而 *gal* 基因的活化子还是 CAP。这样，一种调控蛋白（CAP）与两种不同的抑制子共同作用于不同的基因，这就是一例组合控制（combinatorial control）。实际上，CAP 在 *E. coli* 中与一系列的调控蛋白一起作用于 100 多种基因。

组合控制是基因调控的突出特征。这样，当同一信号控制多组基因时，它一般会以同一调控蛋白将信号传递给每组基因中的一个。这一调控物蛋白也只传递几个信号中参与调控多个基因的那一个。其他信号多数情况下是不同的，则由另外的调控物蛋白介

导。越是复杂的有机体，特别是真核生物中，信号整合的情况就越多。我们以后将看到更复杂，也更精细的关于组合控制的例子（第 17 章）。

选择性的 σ 因子指导 RNA 聚合酶选择性地结合启动子

在第 12 章中我们知道，RNA 聚合酶的 σ 亚基识别启动子序列（图 12-6）。带有 σ^{70} 亚基的 RNA 聚合酶可以识别我们正讨论的 *lac* 启动子和 *E. coli* 的许多其他启动子。*E. coli* 还编码其他 σ 亚基，在特定情况下可以替换 σ^{70}，指导聚合酶选择性地结合启动子。

这些选择因子中有一种热激 σ 因子——σ^{32}。当 *E. coli* 受到热激时，细胞中这种新 σ 因子的量就会升高，它会取代 RNA 聚合酶上的 σ^{70}，指导酶转录那些产物会保护细胞免于受热激影响的基因。σ^{32} 水平的升高由两种机制调控：首先是刺激翻译，即热激后，其 mRNA 以比热激前更大的效率翻译；其二是蛋白质瞬间稳定。另一例选择性 σ 因子，σ^{54} 将在下一部分讨论。σ^{54} 与细胞中小部分的 RNA 聚合酶分子结合指导酶结合到产物参与氮代谢的基因上。

有时候，一系列的选择性 σ 因子指导基因表达的某一特定过程。在细菌枯草杆菌（*B. subtilis*）中发现了两个这样的例子。其中最为精细的部分——该生物孢子形成的控制我们将在第 18 章讨论，其他的部分我们在此简单描述。

噬菌体 SPO1 感染 *B. subtilis*，并在其中生长、溶菌产生后代噬菌体。这一过程要求噬菌体以严格控制的顺序表达其基因。这一控制就是给聚合酶加上一系列的选择性 σ 因子。一旦感染了噬菌体，细菌 RNA 聚合酶（带有 *B. subtilis* 的 σ^{70}）就识别所谓的“早期”噬菌体启动子，指导转录编码感染早期所需蛋白质的基因。其中就有一个基因（称为基因 28）编码一个选择性 σ，它指导聚合酶结合到噬菌体基因组的第二套启动子上，这些启动子与所谓的“中期”启动子相耦联。而这些当中又包括一个编码 σ 因子的基因，该 σ 因子又会指导转录噬菌体“晚期”基因（图 16-14）。

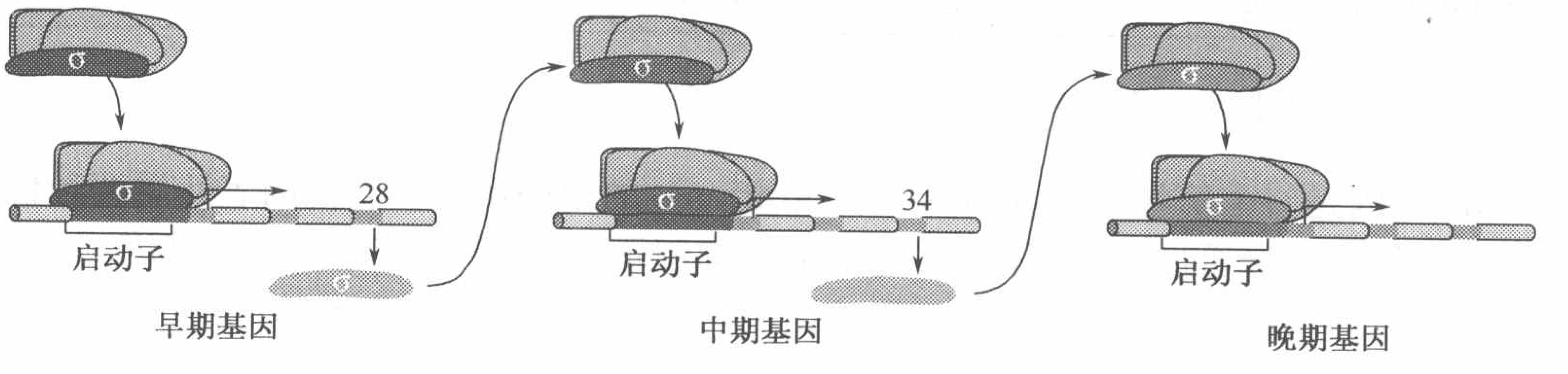


图 16-14 选择性的 σ 因子控制细菌病毒基因的有序表达。细菌噬菌体 SPO1 依次使用 3 个 σ 因子调控其基因组的表达，这保证了病毒基因以所需的顺序表达。（来源：Adapted from Alberts B. et al. 2002. Molecular biology of the cell, 4th edition, p. 415, Fig. 7-63. Copyright © 2002. Reproduced by permission of Routledge/Taylor & Francis Books, Inc.）

NtrC 和 MerR：以变构而非募集作用的转录活化子

大多数活化子都是通过募集起作用的，但是也有例外，NtrC 和 MerR 就是通过变构机制起作用的两例。我们已讲到过变构机制。通过募集起作用的活化子只是将 RNA 聚合

酶的活性形式带到启动子上。而通过变构机制起作用的活化子作用时，聚合酶最初就结合在启动子上，并形成了无活性的复合物，活化子通过引起复合物的变构变化激活转录。

NtrC 控制参与氮代谢的基因的表达，如 *glnA* 基因。RNA 聚合酶预先结合在 *glnA* 基因上，形成稳定的封闭复合体。活化子 NtrC 诱导酶产生构象改变，促使封闭复合体向开放复合体的转变。所以说激活事件就是 RNA 聚合酶的变构化（图 16-2）。

MerR 控制 *merT* 基因，这个基因编码一种酶能够使细胞对汞的毒性效应产生抗性。MerR 也作用于一个无活性的 RNA 聚合酶-启动子复合物。和 NtrC 一样，MerR 也诱导产生构象变化，促使形成开放复合物，但是在这种情况下，活化子的变构效应是施加在 DNA 而非聚合酶上。

NtrC 具有 ATPase 活性且其作用位点距基因较远

如 CAP 相同，NtrC 也具有独立的激活结构域和 DNA 结合结构域并且也只在特异信号存在时才会结合 DNA。对 NtrC 来说，这一信号就是低水平的氮。NtrC 会被激酶 NtrB 磷酸化，并因此而经历构象变化，这一变化使活化子暴露出其 DNA 结合结构域。一旦具有了活性，NtrC 就会结合位于启动子上游 150bp 处的 4 个位点。每一个位点都结合一个 NtrC 二聚体，这些二聚体通过蛋白质-蛋白质相互作用以高度协同的方式结合到 4 个位点上。

转录基因的 RNA 聚合酶的组成包括 σ^{54} 亚基。在没有 NtrC 时，这一聚合酶就结合到 *glnA* 启动子上，形成稳定的封闭复合物。一旦 NtrC（结合到其上游位点）具有了活性，就与 σ^{54} 直接作用。为利于这一相互作用，要求活化子结合位点和启动子之间的 DNA 形成一个环（图 16-15）。如果 NtrC 移动到上游更远的地方（如 1～2kb），活化子仍能起作用。

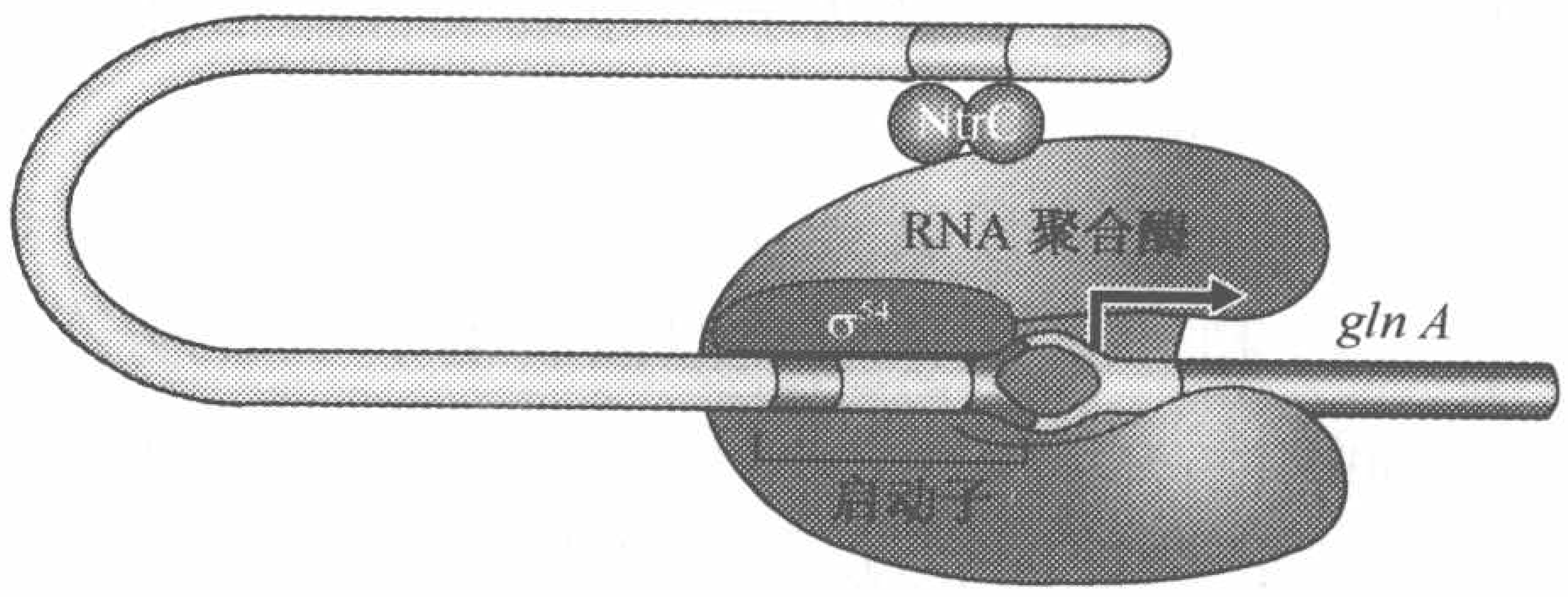


图 16-15　由 NtrC 激活。被包含 σ^{54} 的全酶识别的启动子序列不同于被包含 σ^{70} 的全酶识别的启动子。虽然图中没有特别标示，但 NtrC 与聚合酶的 σ^{54} 亚基接触。NtrC 以二聚体表示，但实际上，它在 DNA 上形成更为有序的复合物。

NtrC 自身也有酶活性——它是一个 ATPase，这一活性可以为其诱导聚合酶产生构象变化提供能量。构象变化促进聚合酶起始转录。具体地说，是刺激稳定的无活性的封闭复合体向活化的开放复合体转化。

NtrC 控制的基因中，有些还具有另一种蛋白质，称为 IHF 的结合位点，位于NtrC 结合位点和启动子之间。IHF 一结合 DNA，就使 DNA 弯曲。当 IHF 结合在正确的位点，DNA 也因而正确弯曲时，就能促进 NtrC 的激活作用。这可以解释为，IHF 通过

弯曲 DNA 使活化子更靠近启动子，协助活化子与结合到启动子的聚合酶相互作用（图 16-4，欲详细了解 IHF 如何弯曲 DNA，见图 11-11）。

MerR 通过扭曲 DNA 激活转录

当 MerR 结合到单一 DNA 结合位点，且有汞存在时，就激活 *merT* 基因。如图 16-16所示，MerR 结合到位于启动子（该基因由包含 σ^{70} 的聚合酶转录）的－10 和－35 序列之间的一段序列上。MerR 结合 DNA 螺旋上 RNA 聚合酶结合点的背面。所以聚合酶可以（也确实是）与 MerR 同时结合到启动子上。

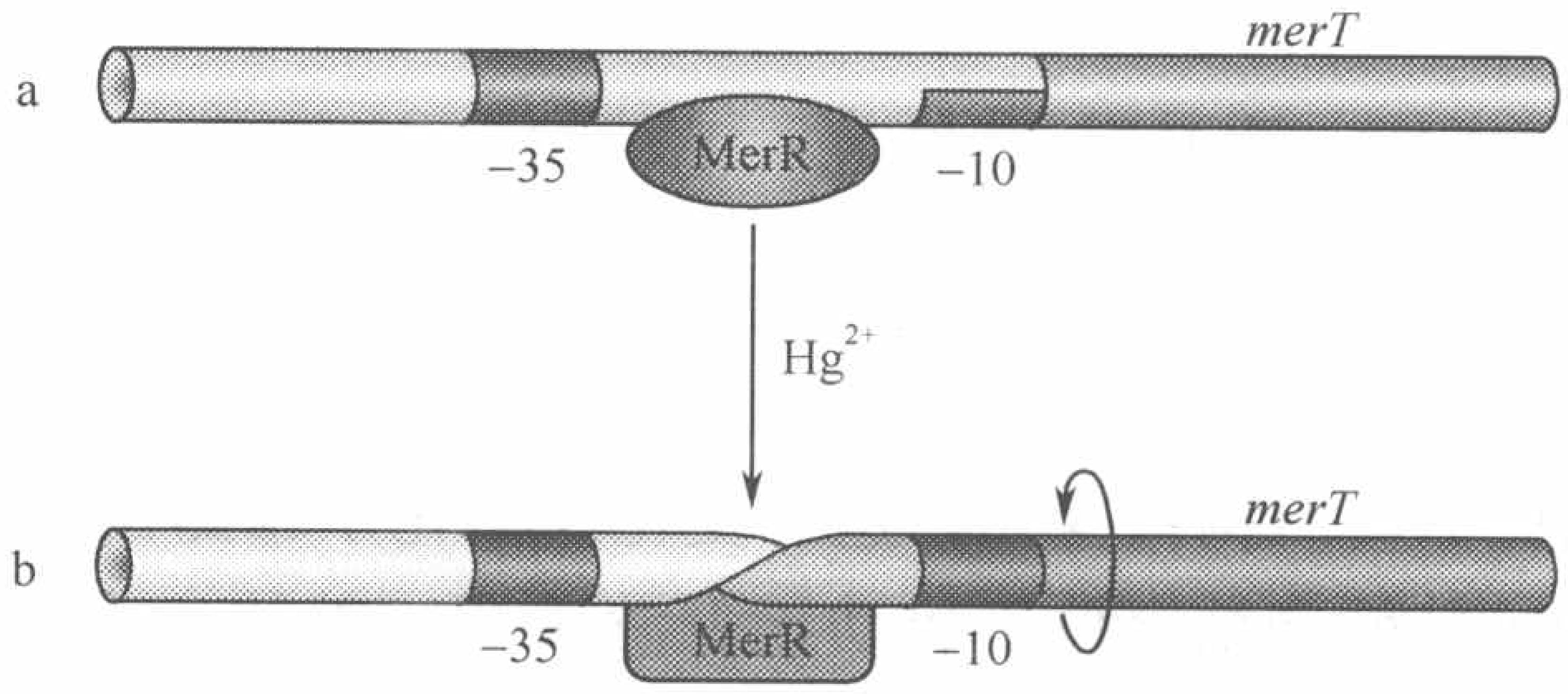


图 16-16　由 MerR 激活。 *merT* 启动子的－10 和－35 组分位于螺旋的反面。a. 在无汞时，MerR 结合并稳定了无活性形式的启动子。b. 在有汞存在时，MerR 扭曲了 DNA 从而使启动子组分正确排列。

merT 不同于一般的启动子，其－10 与－35 组成成分之间的距离是 19bp 而非 σ^{70} 启动子中常见的 15～17bp（第 12 章，框 12-1）。这样将导致这两个被 σ 识别的序列组分既不能适当的间隔也不能平行排列；这两者绕螺旋面彼此相向。而且，MerR 以这种构象在 DNA 上的结合（无 Hg^{2+} 时）锁定了启动子，导致聚合酶虽能结合，但却无法起始转录。因此这种情况下无本底转录。

但是当 MerR 结合了 Hg^{2+} 时，蛋白质就会发生构象变化，引起启动子中心的 DNA 扭曲。这一结构上的弯曲使－10 和－35 区采取类似强启动子 σ^{70} 的分布。在这种新构型中，RNA 聚合酶可以有效地起始转录。启动子 DNA 的“活性”和“非活性”状态的结构已经确定（以另一个同种方式调控的启动子），如图 16-17 所示。

需要着重指出的一点是，在本例中，活化子不是以与 RNA 聚合酶相互作用，而是以改变预先结合的聚合酶附近的 DNA 的构象激活转录的。因此，不像前面的例子，MerR 是没有 DNA 结合区与激活区之分的，其 DNA 结合过程是与其激活过程紧密连接的。

某些抑制子在启动子上滞留而非排斥 RNA 聚合酶

Lac 抑制子以最简单的方式作用：通过结合与启动子有重叠区的位点，阻碍了 RNA 聚合酶的结合。许多抑制子是以这种方式作用的。在 MerR 的实例中，我们看到了一种新的抑制形式。在这一实例中，蛋白质使启动子保持一种不利于转录起始的构象。抑制子还有其他的作用方式，现在我们就来讨论其中的一种。

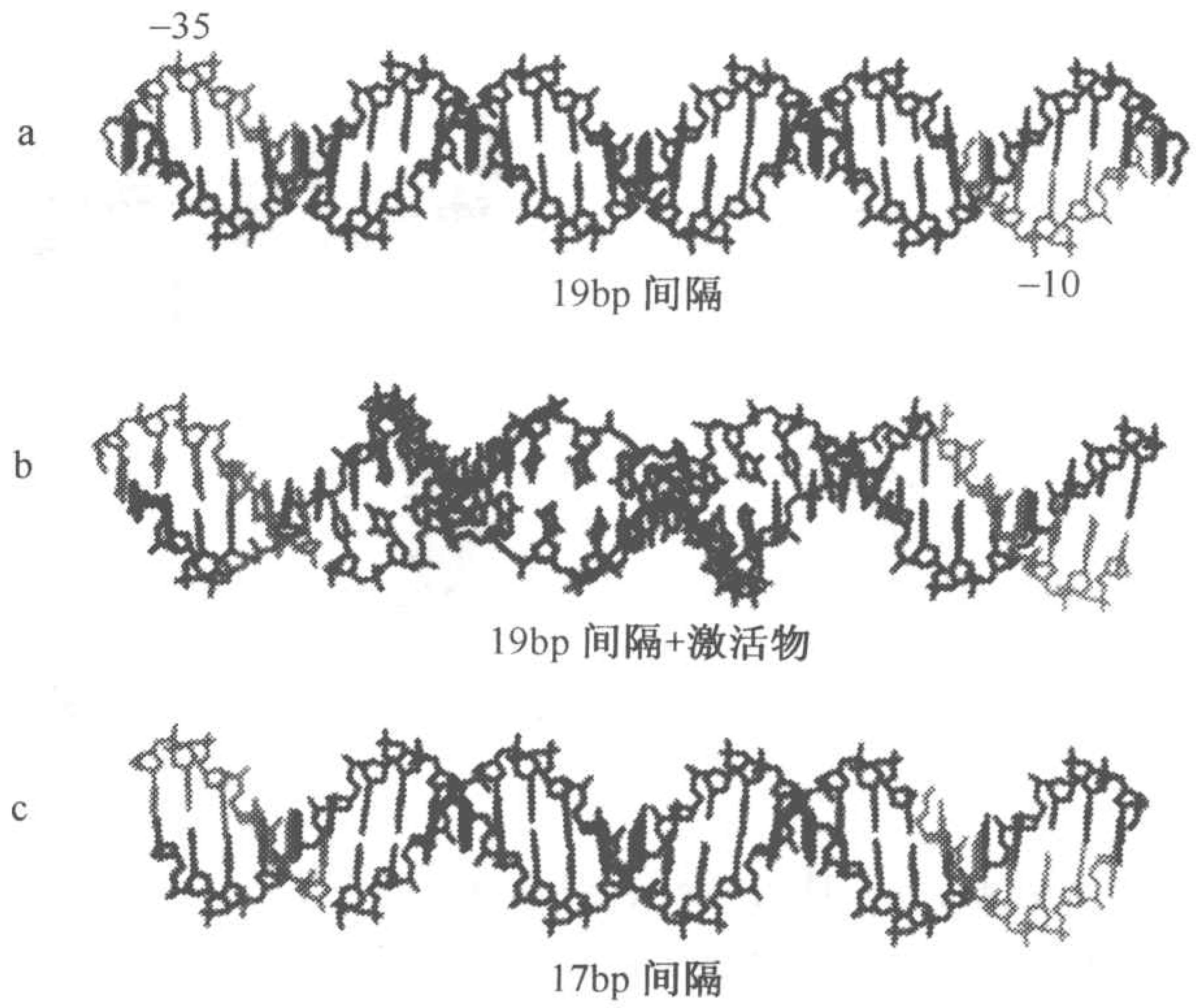


图 16-17　像 *merT* 样的启动子的结构。a. 有 19bp 间隔的启动子。b. 当在带有活性活化子的复合物中时启动子有 19bp 的间隔。c. 有17bp 间隔的启动子。a、b 部分所示的启动子来源于 *Bacillus subtilis* 的 *bmr* 基因，它由调控蛋白 BmrR 控制。BmrR 与 TPP 药物形成复合物就可以作为活化子。−35（TTGACT）和−10（TACAGT）组分的一条链分别以红色和绿色表示。（来源：Adapted，with permission，from Zheleznova Heldwein E. E. and Brennan R. G. 2001. Nature 409：378；Fig. 3 b，c，d. Copyright©2001 Nature Publishing Group.）

有些抑制子的结合位点不与启动子有重叠区。这些启动子不阻碍聚合酶的结合而是结合到启动子旁的位点，与结合在启动子上的聚合酶相互作用，抑制转录起始。我们前面提到的 *E. coli* 的 Gal 抑制子就是一个这样的例子。Gal 抑制子控制编码参与半乳糖代谢的酶的基因。在没有半乳糖的时候，抑制子保持基因不表达。在这种情况下，抑制子与聚合酶相互作用，抑制封闭复合物到开放复合物的转变。

另一个例子来自在细菌 *B. subtilis* 的噬菌体（ϕ29）的 P_4 蛋白。这一调控蛋白结合到一个启动子——一个弱化的启动子 P_{A3} 的附近，与聚合酶相互作用作为一个活化子，它与 α CTD 相互作用，与 CAP 相同。但这个活化子还结合另一个启动子——一个强启动子称为 P_{A2c}。与在弱启动子上一样，该调控蛋白也与聚合酶接触，但在这里却是导致抑制。在前一种情况下额外的结合能协助募集聚合酶，因此激活基因；而在后一种情况下，由聚合酶和启动子之间的强相互作用加上活化子相互作用所提供的总结合能太强，以至于聚合酶无法从启动子上脱逃。

AraC 和抗激活作用对 *araBAD* 操纵子的控制

E. coli 的 *araBAD* 操纵子的启动子在阿拉伯糖存在而没有葡萄糖时被激活，指导编码阿拉伯糖代谢需要的酶的基因表达。有两个活化子一起作用：AraC 和 CAP。当存在阿拉伯糖时，AraC 结合这种糖，采取允许其结合 DNA 的构象，它以二聚体结合到紧邻的两个半位点，*araI*$_1$ 和 *araI*$_2$（图 16-18a）。在这两个位点的上游就是一个 CAP 位点（图中未标示）：无葡萄糖时，CAP 结合到这里协助激活转录。

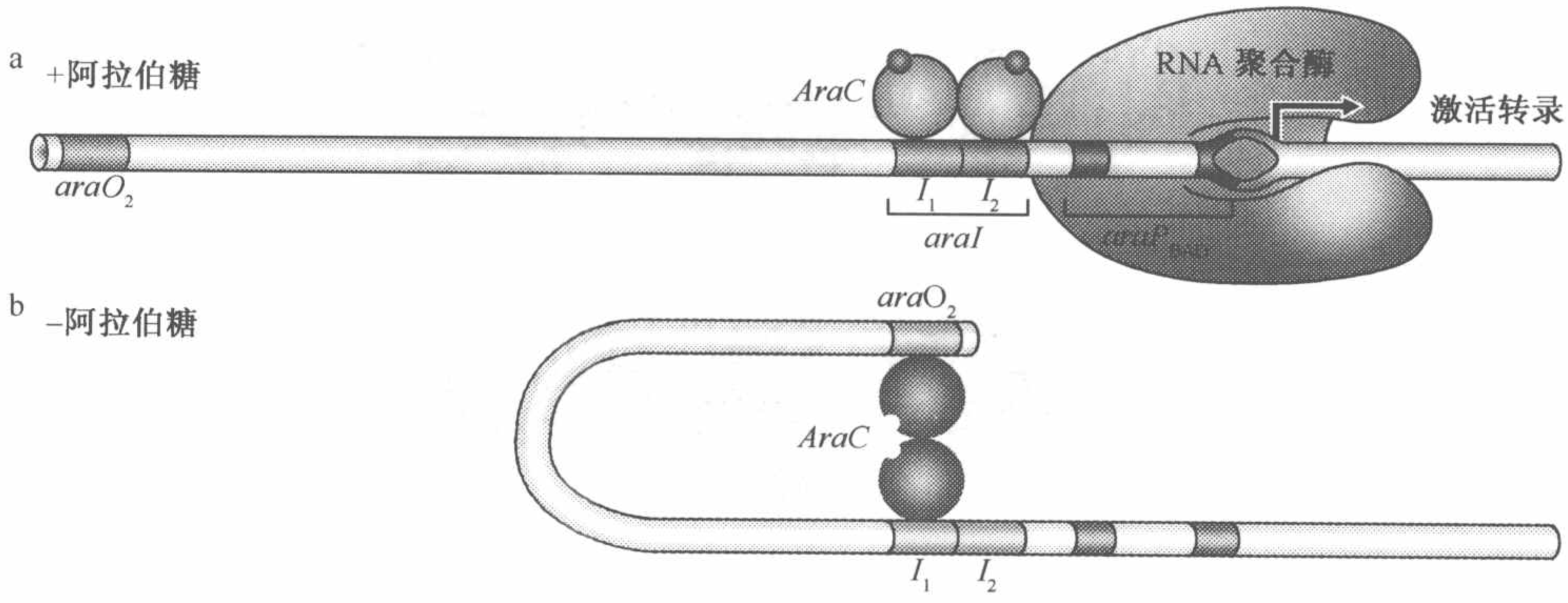


图 16-18 *araBAD* 操纵子的控制。 a. 阿拉伯糖结合到 AraC，改变了活化子的形状，使它以二聚体结合到 $araI_1$ 和 $araI_2$ 上，这将使一个 AraC 单体置于启动子附近，它可以从此处激活转录。b. 在无阿拉伯糖时，AraC 二聚体采取一种不同的构象，结合到 $araO_2$ 和 $araI_1$。在这一位置，$araI_2$ 位点上没有单体，因此蛋白质不能激活 *araBAD* 启动子。这一启动子也被 CAP 控制，但图中没有标示。

没有阿拉伯糖时，*araBAD* 基因不表达。这是因为，AraC 在不结合阿拉伯糖时，会采取不同的构象以不同的方式结合 DNA：一个单体仍结合 $araI_1$ 位点，但另一个单体则结合远处一个叫 $araO_2$ 的半位点，如图 16-18b 所示。因为这两个半位点相距 194bp，所以当 AraC 以这种方式结合上去时，两个位点之间的 DNA 就会形成环。同时，这样结合，$araI_2$ 上没有 AraC 结合，而启动子 *araBAD* 又是从这一位置被介导激活，所以，这样的构型不会激活启动子。

阿拉伯糖诱导 *araBAD* 启动子的数量级是非常大的，因此，该启动子经常被用作表达载体（expression vector）。表达载体是构建的 DNA。某一蛋白质其基因与强启动子融合就可在表达载体中有效合成（第 20 章）。以某一基因与 *araBAD* 融合就可以使这一基因的表达由阿拉伯糖来控制：该基因可以一直处于关闭状态除非想要其表达，如果想要其产物，只需加入阿拉伯糖诱导即可。这样，即使其产物对细菌细胞具有毒性，表达也被允许。

λ 噬菌体：调控的层次

细菌噬菌体 λ 是一种感染 *E. coli* 的病毒。它一旦感染细菌，就会以两种方式繁殖：裂解（lytically）或溶原生长（lysogenically），如图 16-19 所示。裂解生长需要噬菌体 DNA 的复制和新的外壳蛋白的合成。这些组分共同形成新的噬菌体颗粒，通过宿主细胞的裂解释放出来。溶原现象——一种选择性繁殖途径，包括噬菌体 DNA 整合进入细菌染色体并像细菌基因组的正常组分似的在每次细胞分裂的时候主动复制。

溶原性细菌在正常情况下非常稳定，但其体内处于静止的噬菌体［原噬菌体（prophage）］在细胞 DNA 遭受破坏时就会有效地转向裂解生长（因此会威胁到宿主细胞的继续生存）。这种从溶原性生长到裂解生长的转变称为溶原性诱导（lysogenic induction）。

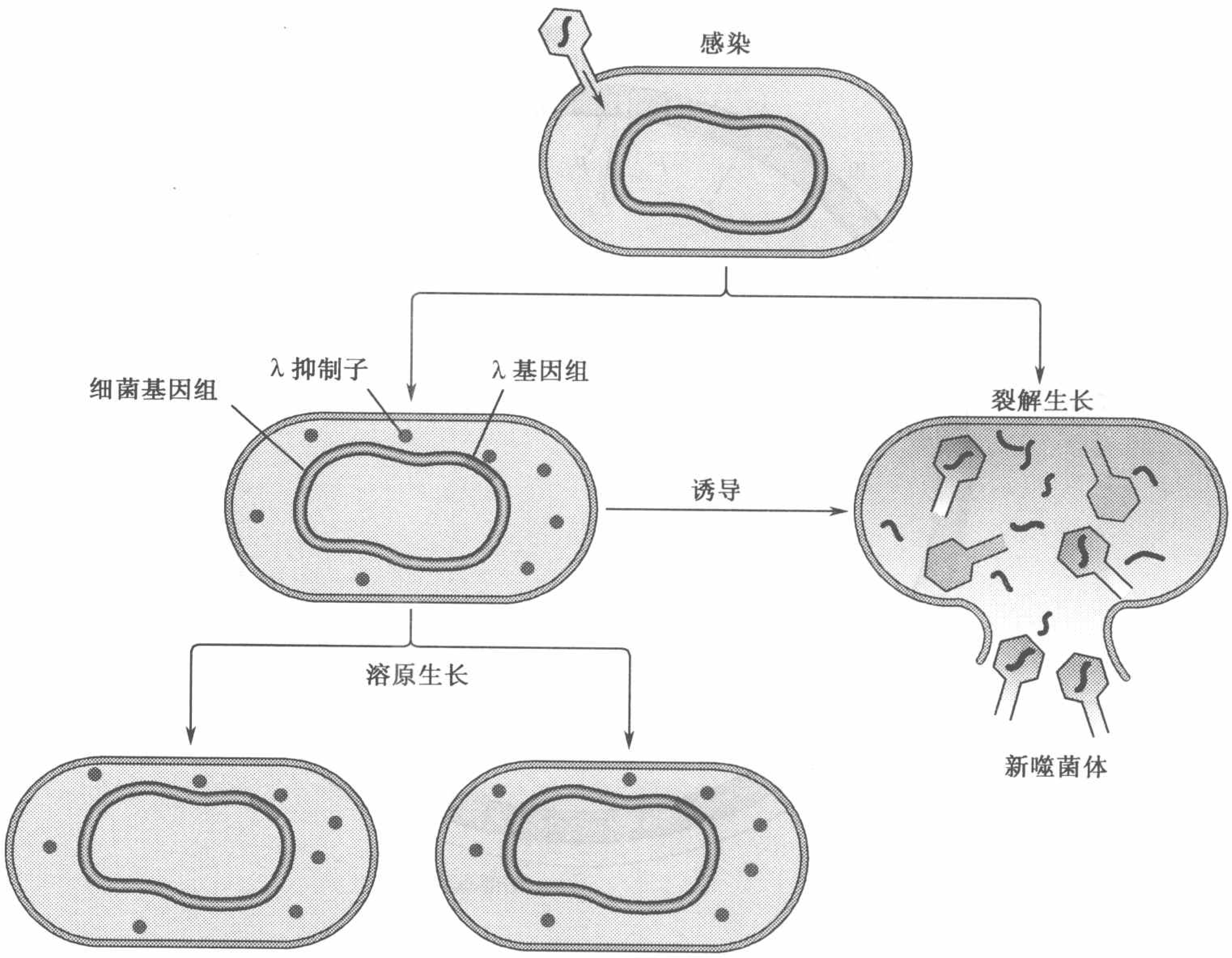


图 16-19 λ溶原体的生长和诱导。一旦感染，λ就可以进行裂解或溶原过程。溶原体可以稳定的增殖许多代，它也可以被诱导。诱导之后，裂解基因以正确的顺序表达，导致新的噬菌体颗粒的产生。

对发育途径的选择依赖于细胞对两种选择性的基因表达程序的选择。溶原状态的表达程序可以维持许多代，但是，一旦诱导，就会有效地转向裂解程序。

基因表达的选择性模式控制裂解和溶原性生长

λ噬菌体有 50kb 的基因组和大约 50 个基因，其中的大多数编码外壳蛋白，参与 DNA 复制的蛋白、重组蛋白和裂解蛋白（图 16-20）。这些基因的产物对于在裂解周期中形成新的噬菌体颗粒非常重要，但在此我们关心的是那些调控蛋白及其作用。因此，我们可以集中于这些基因中的几个。我们先来考虑基因组中一个很小的区域，如图 16-21所示。

图中描述的区域包括 2 个基因（*cI* 和 *cro*）和 3 个启动子（P_R、P_L 和 P_{RM}）。所有其他的噬菌体基因（除去一个小基因）都在该区之外直接从 P_R 和 P_L（分别表示右向和左向启动子）转录，或从其他启动子起始转录，这些启动子的活性由从 P_R 和 P_L 开始转录的基因的产物控制。P_{RM}（promoter for repressor maintenance）只转录 *cI* 基因。P_R 和 P_L 是强组成型启动子，即它们有效结合 RNA 聚合酶并不需活化子的协助而指导转录。相反，P_{RM}是一个弱启动子且只在上游结合活化子后才会有效指导转录。在这方面，P_{RM}类似于 *lac* 启动子。

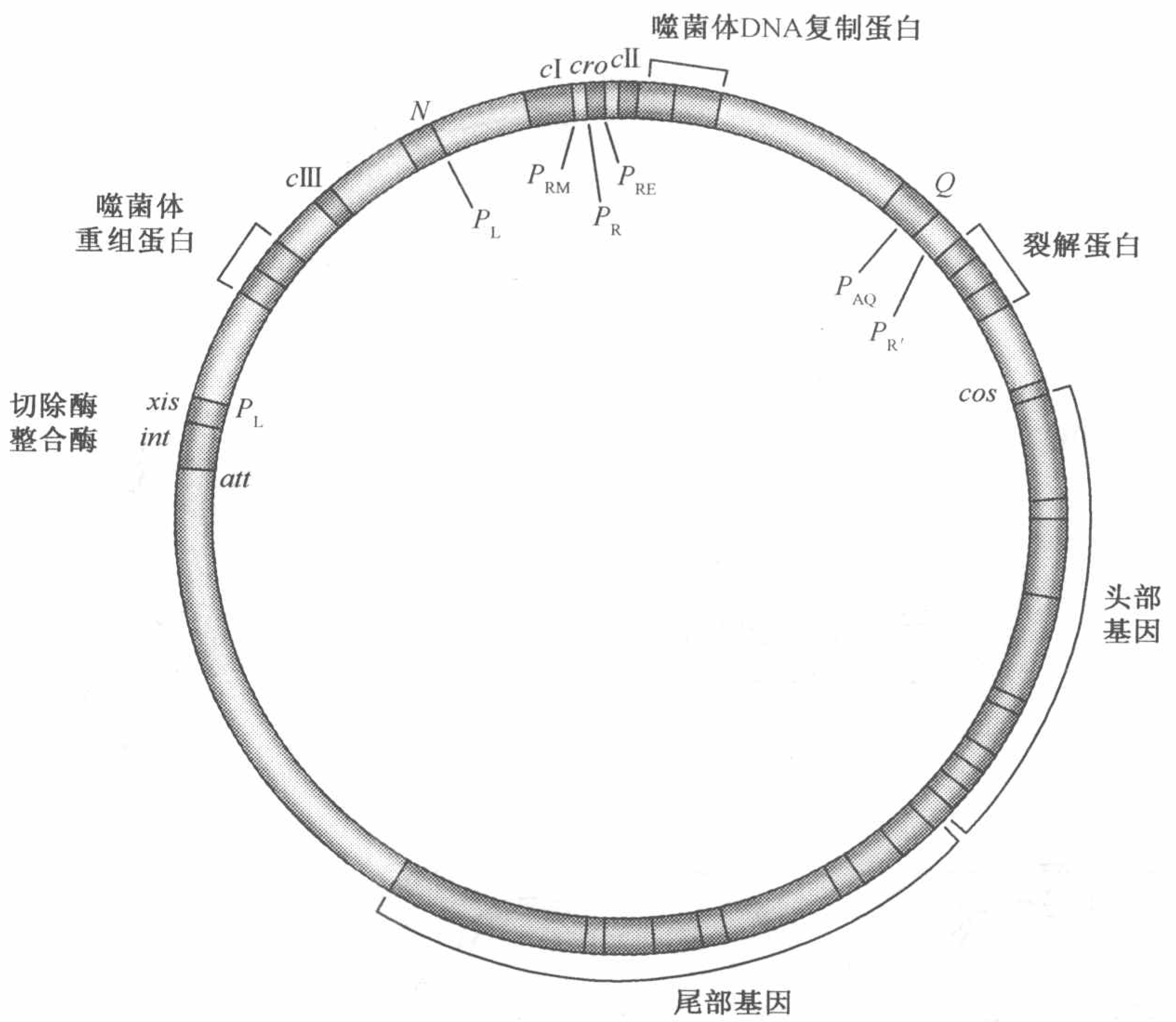


图 16-20　噬菌体 λ 环形图。λ 基因组在噬菌体头部是线形的，但一旦感染，就会在 *cos* 位点环化。当整合进细菌染色体时，它又会线形化，末端在 *att* 位点。（见第 11 章关于整合的描述）

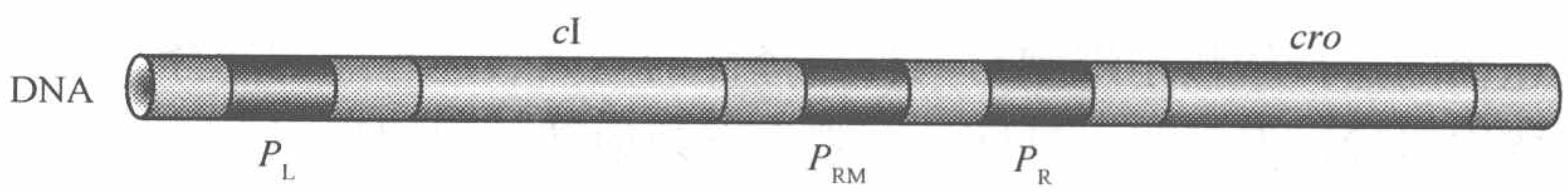


图 16-21　λ 噬菌体左侧和右侧的启动子控制区域。

图 16-22 中所示的基因表达有两种顺序：一种出现在裂解生长周期，另一种出现在溶原生长周期。当 P_R 和 P_L 持续开放而 P_{RM} 关闭时，就发生裂解生长。相反，溶原生长则是 P_R 和 P_L 关闭而 P_{RM} 打开的结果。这些启动子是如何控制的呢？

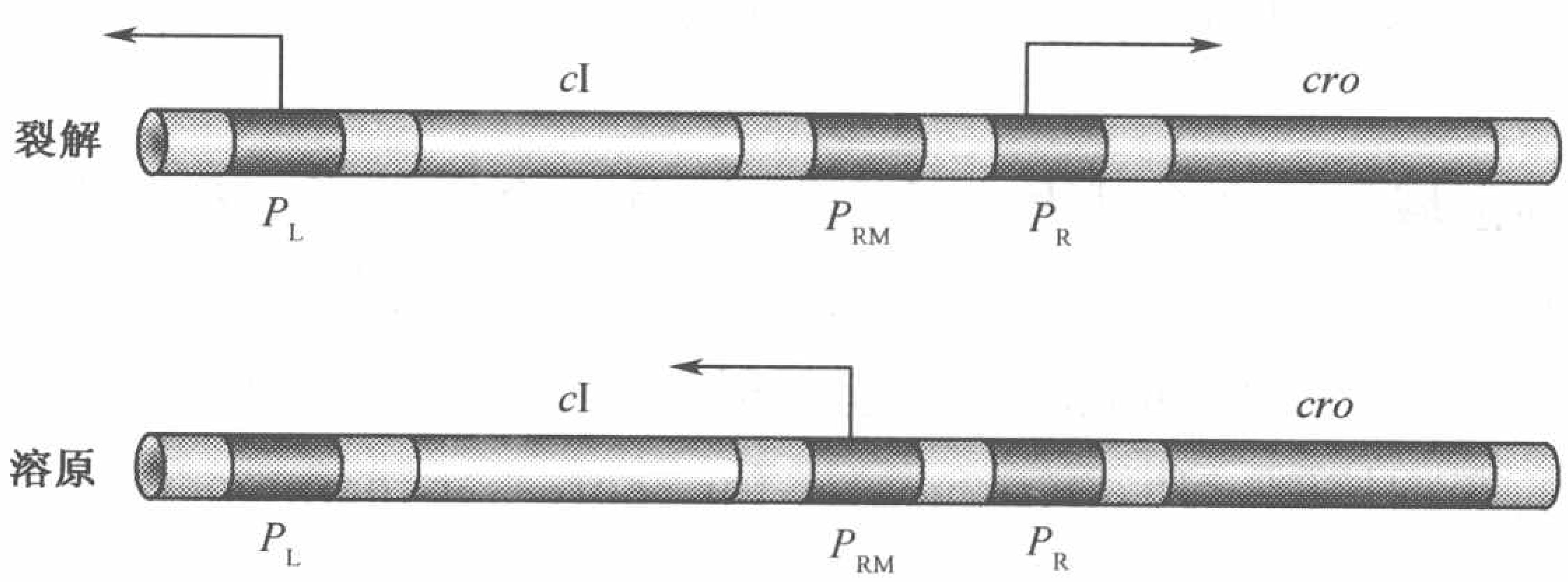


图 16-22　裂解和溶原生长中 λ 控制区域的转录。箭头分别指示在裂解和溶原生长决定期哪一个启动子是有活性的。箭头也指示了从每一个启动子开始转录的方向。

调控蛋白及其结合位点

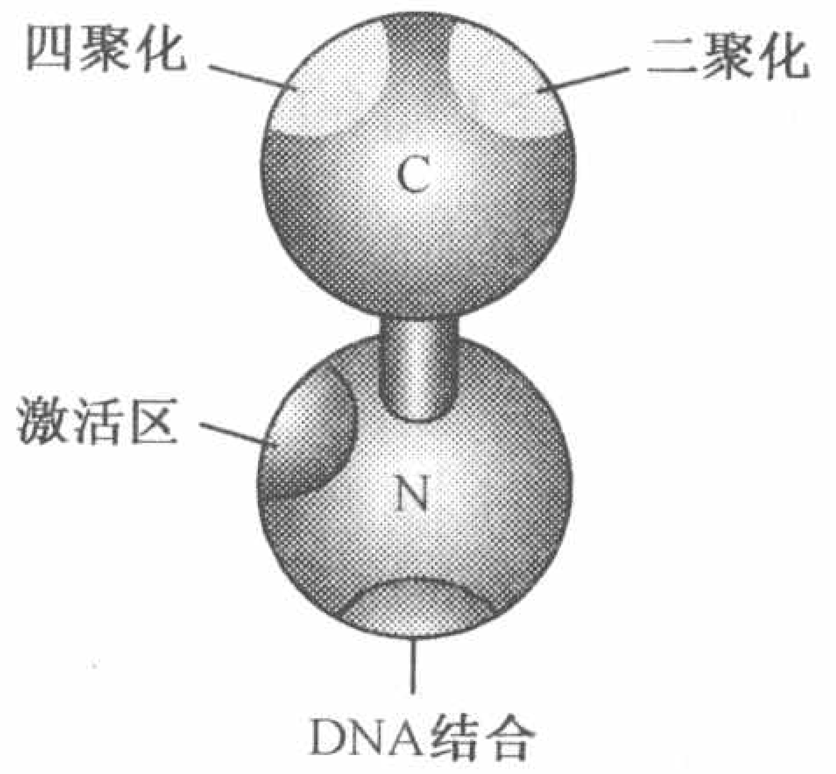


图 16-23 λ 阻抑物。图中示 λ 阻抑物的一个单体，表示了参与该蛋白质不同活性的各种表面。N 表示氨端结构域，C 表示碳端结构域，“四聚化”表示两个二聚体协同结合到相邻的 DNA 位点时，相互作用的区域。（来源：Adapted from Ptashne M. and Gann A. 2002. Genes & signals, p. 36, Fig. 1～17. © Cold Spring Harbor Laboratory Press.）

cI 基因编码 λ 抑制子。λ 抑制子是一种包含由柔性衔接区连接的两个结构域的蛋白质（图 16-23）。N 端结构域包括 DNA 结合区（一个螺旋-转折-螺旋结构域，与我们以前所见的相同）。与 DNA 结合蛋白质一样，λ 抑制子也是以二聚体形式结合 DNA 的；二聚化时，两个单体的 C 端结构域接触。一个二聚体识别一段 17bp 的 DNA 序列，每一个单体识别一个半位点，这也与我们在 *lac* 系统所见是相同的（我们已经在图 16-12 中详细地看到过 λ 抑制子对 DNA 的识别）。

尽管其名为“抑制子”，但 λ 抑制子既可以激活也可以抑制转录。当作为抑制子起作用时，它与 Lac 抑制子以相同的作用方式结合到与启动子有重叠区的位点排斥 RNA 聚合酶。作为一个活化子时，λ 抑制子就类似 CAP，通过募集起作用。λ 抑制子的激活区位于蛋白质的 N 端结构域，它在聚合酶上的靶标是 σ 亚基上的一个区域，该区邻近于 σ 的启动子—35 识别区域（4 区，第 12 章，图 12-6）。

Cro（意为抑制子和其他基因的控制，control of repressor and other thing）只抑制转录，类似于 Lac 抑制子。它是一个单结构域蛋白并且也是以二聚体形式结合到 17bp 的 DNA 序列。

λ 抑制子和 Cro 可以结合 6 个操作子中的任意一个，每个蛋白质以不同的亲和力识别这些位点，3 个位点位于左控制区，3 个位于右控制区。我们将重点讨论 λ 抑制子和 Cro 在右侧区位点的结合，如图 16-24 所示。左侧区位点的结合模式与右侧区类似。

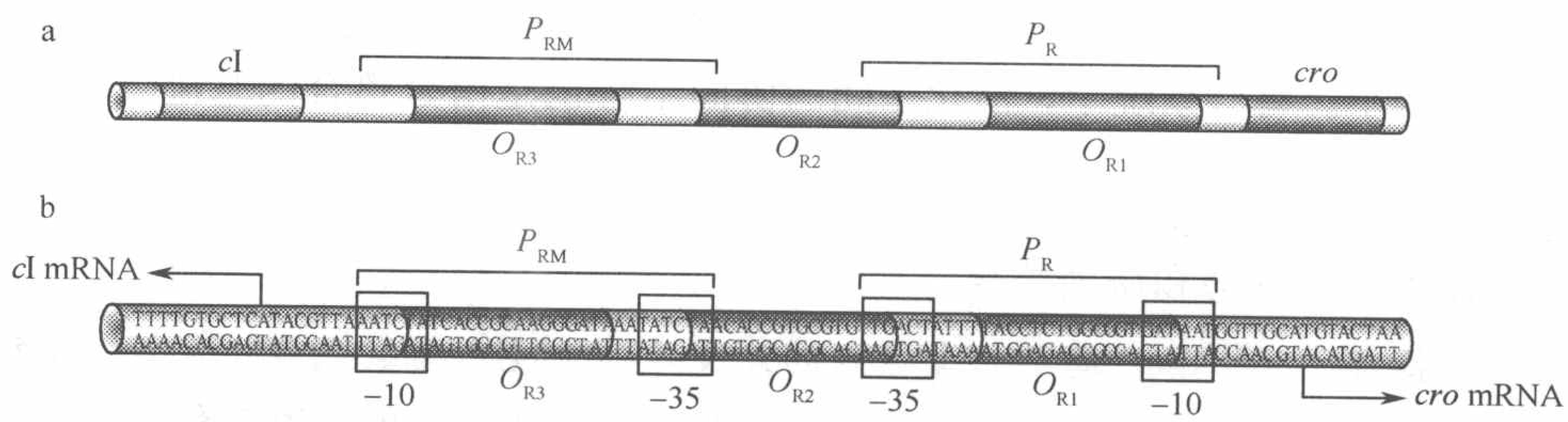


图 16-24 O_R 中的启动子和操作子位点的相关位置。注意 O_{R2} 和 P_R 的—35 区有 3 个碱基对的重叠，与 P_{RM} 有 2 个碱基的重叠。这一点不同就足以使 P_R 被结合到 O_{R2} 上的抑制子抑制，同时 P_{RM} 被激活。（来源：b. Adapted from Ptashne M. 1992. A genetic switch: Phage and higher organisms, 2 nd edition. Copyright©1992. Blackwell Science Ltd. Used with permission.）

位于右侧操作子的 3 个位点称为 O_{R1}、O_{R2} 和 O_{R3}；这些位点的序列相似，但是并不完全相同。每一个位点（如果与其他位点分离开来并分别检测的话）都可以结合一个抑制子或者 Cro 二聚体。但是其相互作用的亲和力是不一样的。抑制子结合 O_{R1} 是它结合

O_{R2}的亲和力的 10 倍。换句话说，结合 O_{R2}所需的抑制子的量是结合 O_{R1}的 10 倍或者说浓度是结合 O_{R1}的 10 倍，O_{R3}结合抑制子的亲和力与 O_{R2}的相同。Cro 则相反，它结合的亲和力最强，而结合 O_{R1}和 O_{R2}则需 10 倍浓度。这些亲和力差异的意义在下文很快就会明了。

λ 抑制子与操作子位点协同结合

λ 抑制子协同结合 DNA，这对其功能是至关重要的，结合的发生如下。考虑抑制子结合 O_R 位点。λ 抑制子的 C 端结构域不仅提供二聚化时的接触，还介导二聚体间的相互作用（接触点即是图 16-23 所示的“四聚化”补丁区），这样，两个抑制子二聚体就可以协同结合到 DNA 上临近的位点。

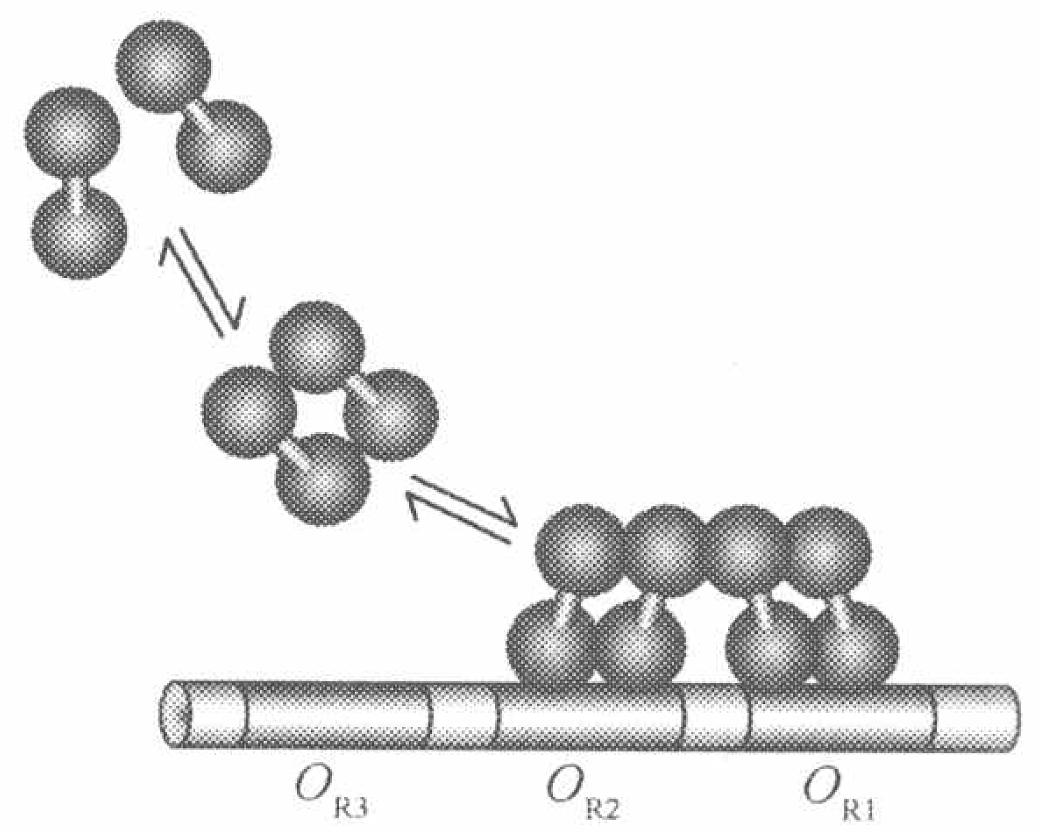


图 16-25 λ 抑制子协同结合 DNA。λ 抑制子单体相互作用形成二聚体，二聚体相互作用形成四聚体，这些相互作用保证了抑制子结合 DNA 是协同的。在 O_R 和 O_L 上的四聚体的相互作用增强了这种协同结合。（见下文及图 16-27）

例如，O_{R1}上结合的抑制子通过协同结合协助抑制子结合到低亲和力位点 O_{R2}。这样抑制子就可以自动结合两个位点，并且在浓度只够单独结合 O_{R1}时就可以做到（图 16-25）。（回顾所学内容，在无协同性时，结合 O_{R2}需要的抑制子浓度要高 10 倍。）O_{R3}不被结合：抑制子协同结合 O_{R1}和 O_{R2}但却不能自动地与第三个二聚体在这一邻近位点结合。

我们前面已经讨论过协同结合的观点并看了一个例子：CAP 激活 *lac* 基因。在例子中，协同结合是抑制子自动结合到同一 DNA 分子上的位点时相互接触的直接结果。

框 16-3 详细讨论了协同结合的原因和影响。调控蛋白的协同结合是为了保证基因的控制信号水平小的变化也可以使基因的表达水平有很大改变。下面讨论的 λ 溶原性诱导就提供了一个很好的控制灵敏性方面的例子。有些系统中，活化子间的协同结合也是信号整合的基础（见第 17 章 β-干扰素的讨论）。

框 16-3 浓度、亲和力和协同结合

我们所提及“强”和“弱”结合位点的意义是什么呢？当我们说，2 个分子彼此识别或相互作用（如蛋白质和 DNA 上的位点）是指它们彼此存在某种亲和力。它们是否能在某一特定时间结合在一起取决于两个方面：① 亲和力有多高，如它们相互作用有多紧密；② 分子的浓度。

如我们在第 3 章和第 5 章强调的，支撑生物系统调控的分子间相互作用是可逆的：当两个可以相互作用的分子相遇时，它们就黏附在一起，一段时间以后又会分开。亲和力越高，两个分子黏附得越紧密，一般来说，它们在一起的时间也就越长。浓度越高，它们在第一时间彼此相遇的概率就越高。因此，高亲和力和高浓度的效应是近似的：一般两种情况都导致两个分子结合在一起的时间延长。

协同性的视觉展示

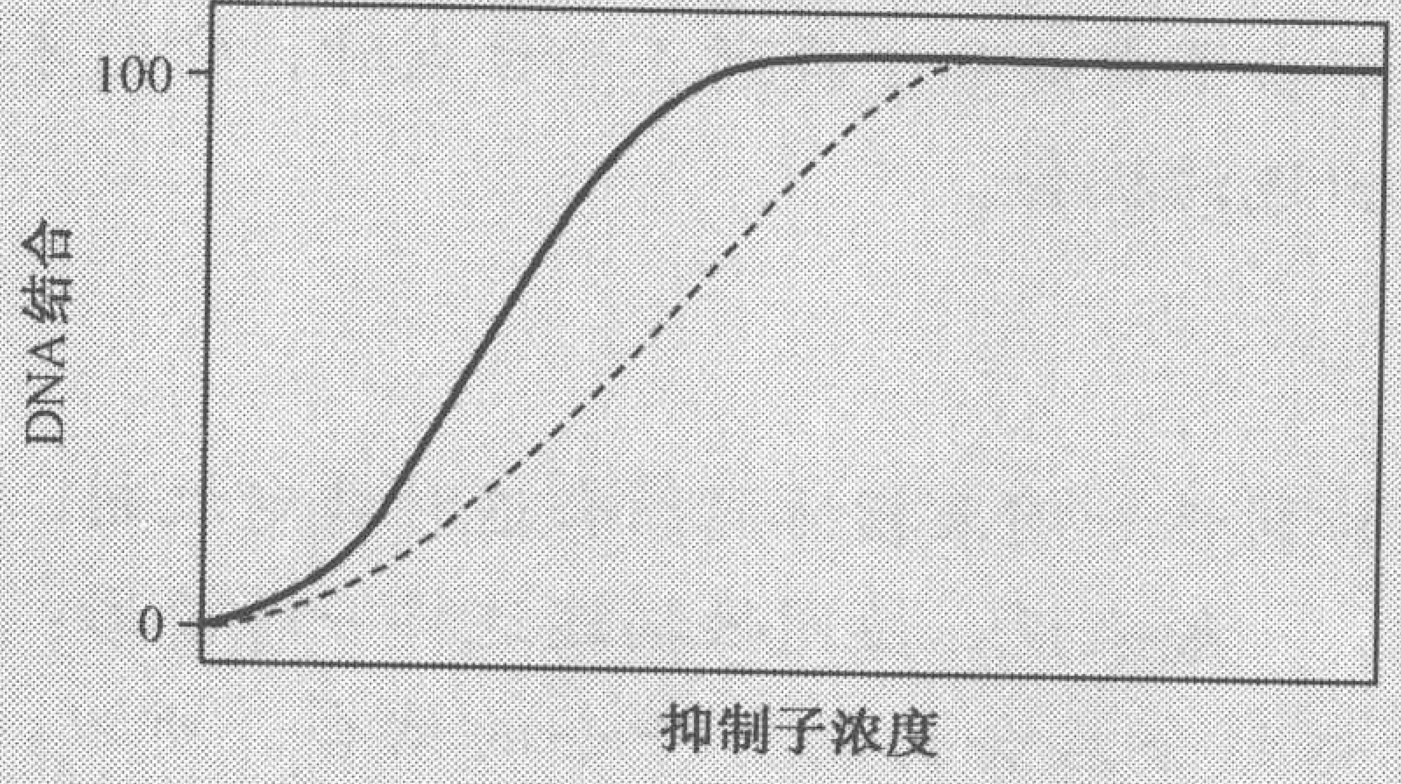
协同性表现为亲和力的提高。抑制子对O_{R1}的亲和力比对O_{R2}要高。但是一旦抑制子结合了O_{R1}，结合O_{R2}的抑制子就会结合得更加紧密，因为它不仅与O_{R2}作用还与结合在O_{R1}上的抑制子相互作用，这两个相互作用各自的亲和力并不很强，但二者联合时则大大提高了第二个抑制子的亲和力。我们在第 4 章时看到，结合能和平衡之间是指数关系（表 4-1）。因此，即使仅将结合能提高 2 倍，亲和力也能大规模提高。

另一个描绘协同性如何工作的方式是将其认定为抑制子浓度的局部升高。将协同结合在O_{R1}和O_{R2}的抑制子放在一起。尽管O_{R2}上的抑制子会周期性地释放 DNA，但它却会保留在结合于O_{R1}的抑制子上，因而与O_{R2}保持靠近。这将有效地提高位点附近的抑制子的浓度，保证抑制子频繁的重新结合上去。

假设无协同性而只是提高细胞内抑制子的浓度，那么当抑制子从O_{R2}上脱落后，它不会被结合在O_{R1}上的抑制子存留在附近，而是通常在结合O_{R2}之前就游离开。但是在抑制子有较高的浓度时，另一抑制子分子有可能位于O_{R2}附近并结合上去。因此，即使每一个抑制子二聚体只在O_{R2}上呆一段较短时间，也能通过将它保持在附近或者提高置换的可能次数，而在特定时间内提高结合的概率。

还有一种考虑协同结合的方法，即考虑熵的影响。当蛋白质从溶液中的自由状态转变为束缚在 DNA 结合位点上时，系统的熵会降低。而位于O_{R2}附近的抑制子受到O_{R1}上抑制子的相互作用，与自由状态下相比已被束缚。束缚的抑制子比自由抑制子结合 DNA 的熵值小。

这样，我们看到了 3 种描绘协同性的方式。我们还应该考虑协同结合的结果，协同结合在生物学中是很有用的。例如，协同性不仅能使一个弱结合位点在蛋白质浓度低于其自身亲和力时被其饱和，还能改变描述位点结合蛋白的饱和度随蛋白质浓度改变的曲线的坡度。为了理解这一点，我们以一个协同结合到两个弱位点 A 和 B 的蛋白质为例。这两个位点从原本完全无结合蛋白到几乎完全饱和所需的蛋白质浓度的变化范围比单位点时窄很多（框 16-3 图 1）。实际上，λ系统中的协同性比想像的要大，因为在细胞中发现有大量的单体形式的自由抑制子（如未结合 DNA 的抑制子），因此，实质上，结合的时候是 4 个单体而不是两个稳定的二聚体协同结合，这增加了在 DNA 上形成复合物时的协调性，也就增加了曲线的坡度。但是为什么协同性会使结合曲线坡度更大呢？



框 16-3 图 1　协同性结合反应。点线表示蛋白质与一个位点的结合曲线，陡的 S 型曲线表示是像λ抑制子与自己的操作子位点结合。（经允许引自 Ptashne M. 1992. A genetic swith: Phage and higher organisms, 2 nd ed. 版权所有 Blackwell Science.）

我们已经看到结合位点是如何在抑制子浓度低于亲和力要求时被饱和的。但是，为什么当抑制子浓度降低时，结合也会降低得这么快呢？让我们考虑任一系统的各组分间的相互作用：当降低组分的浓度时，任意两个组分的相互作用发生的概率将减小。如果该系统要求几个不同组分间的多重相互作用，那么在低浓度很少会发生相互作用。而蛋白质的 4 个单体结合到 2 个位点上时，就需要几个（实际是 7 个）相互作用；当它们每一个的浓度降低时，个别组分相遇的概率就大大降低了。

协同性和 DNA 结合特异性

最后，协同结合还有一个重要的意义是，它赋予 DNA 结合特异性。CAP 激活 *lac* 启动子表明了这一点。CAP 特异的将 RNA 聚合酶带到带有 CAP 位点的启动子上（而不是其他无 CAP 但亲和力相当的位点）。同样地，λ 抑制子在 O_{R1} 处指导另一抑制子分子结合到它邻近的弱位点，而不是细胞中其他具有相同亲和力的地方。实际上，协同性对于保证蛋白质以足够的特异性结合从而使生命正常活动是至关重要的。

为了阐明这一点，我们来考虑一个蛋白质结合 DNA 位点的情况。这一蛋白质对于其正确位点具有高亲和力。但细胞内的 DNA 却意味着大量潜在（但不正确）的蛋白质结合位点。因此，重要的是，该蛋白质不仅要对其正确的位点有绝对的亲和力，而且要有相对于所有其他非正确位点的亲和力。请注意，那些非正确位点（代表着除正确位点外的所有 DNA）的浓度比正确位点的浓度高得多。所以，即使非正确位点的亲和力低于正确位点，前者的较高浓度也会保证蛋白质在试图结合正确位点的时候经常随机结合这些非正确位点。

这就需要一种策略来提高对正确位点的亲和力却不至于提高对非正确位点的亲和力。增加蛋白质和 DNA 位点间的接触（如合成更大的蛋白质）不一定有用，因为这同时也增强了蛋白质与非正确位点的结合。一旦蛋白质对非正确位点的亲和力太高，那么它基本上就无法找到正确位点；随机结合非正确位点占去了它太长的时间。这样，动力学问题取代了特异性问题，而这同样不利于特异性结合。

协同性解决了这一问题，蛋白质通过协同结合到两个邻近的位点，大大提高了它对这两个位点的亲和力，却不增强对其他位点的亲和力。协同性不会增强对非正确位点亲和力的原因是，2 个蛋白质分子同时结合（允许协同性稳定这种结合）2 个靠近的非正确位点的机会极小。只有当蛋白质分子找到正确位点时，它们才会保持足够长的结合时间，给予第二个蛋白质出现的机会。

协同性和变构效应

尽管本章中我们用协同性一词来指协同结合这种特定机制，但这个词还被用在其他环境中代表不同的机制。一般来说，协同性描述这样一种情形：两个配体结合第三个分子，两个配体中一个配体结合后会协助另一个的结合。所以，对我们在这里考虑的 DNA 结合蛋白来说，协同性是由简单的黏附相互作用介导的，而在别的情形下，则可能是由变构事件介导的，氧分子结合血红素可能是最好的例子。

血红素是同源四聚体，每一个亚基结合1分子的O_2。这种结合是协同的：当第一个O_2结合后，它引起构象改变，修饰了第二个结合位点，使其采取高亲和力构象。所以，在这一个例子中，没有配体间的直接相互作用，而是引起变构转变，一个配体增强了对第二个配体的亲和力。

抑制子和Cro以不同模式结合控制裂解和溶原生长

抑制子和Cro是如何控制与不同的λ生长方式耦联的不同基因的表达模式的呢？对裂解生长来说，Cro二聚体结合O_{R3}；这一位点与P_{RM}重叠，因此，Cro抑制了该启动子（由于这一启动子是弱启动子，所以在无活化子时只以低水平作用，图16-26）。因为抑制子和Cro都不结合O_{R1}和O_{R2}，所以P_R结合RNA聚合酶并指导裂解基因的转录；P_L作用方式与此相同。请回顾前文所述，P_R和P_L都是无需活化子的强启动子。

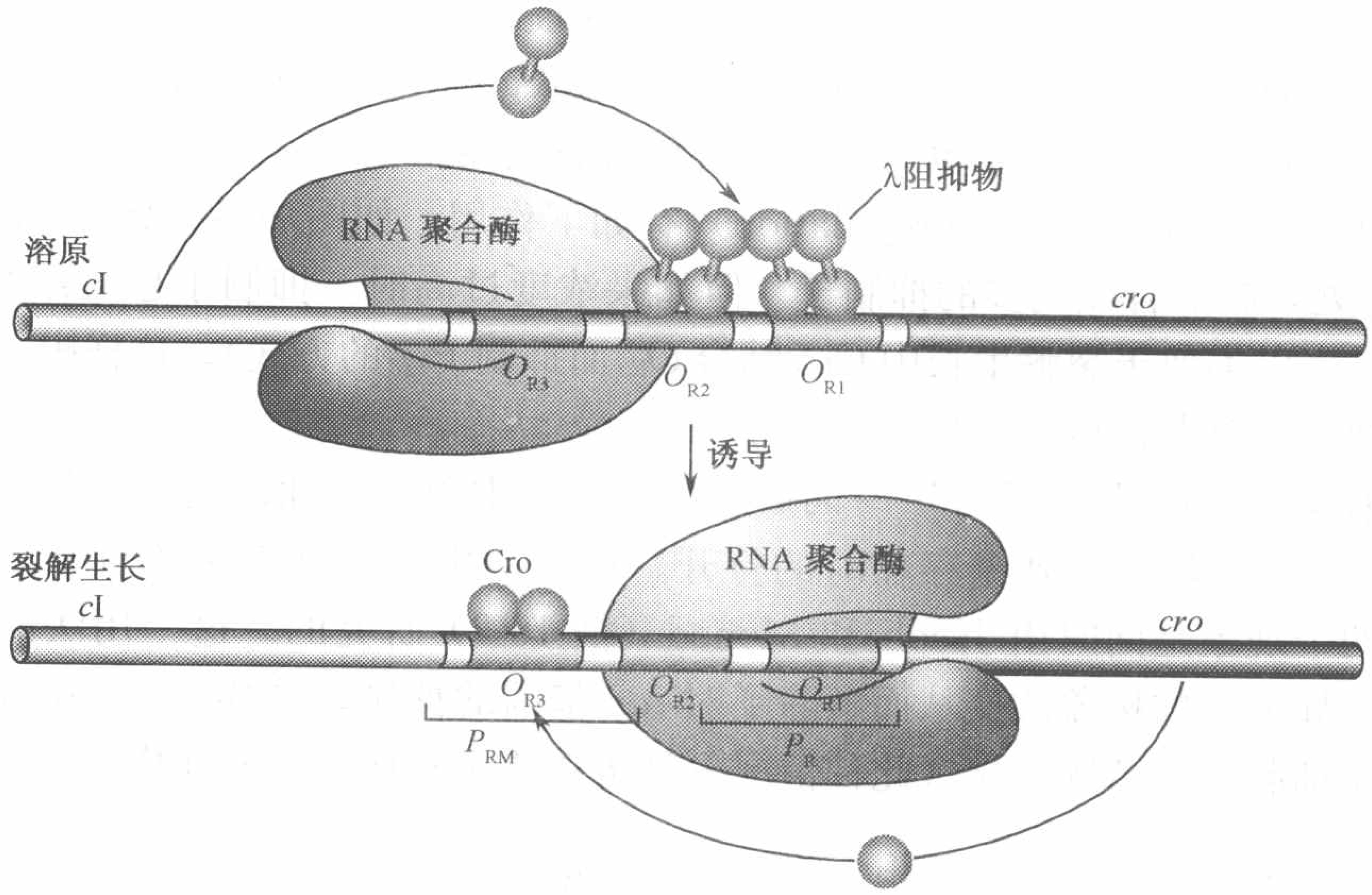


图16-26 λ抑制子和Cro的活性。抑制子结合到O_{R1}和O_{R2}上关闭了从P_R开始的转录。结合到O_{R2}的抑制子P_{RM}上的RNA聚合酶接触，激活了*c*I（抑制子）基因的表达。O_{R3}在P_{RM}内，结合在那里的Cro抑制了*c*I的转录。（来源：Adapted from Ptashne M. and Gann A. 2002. Genes & signals, p. 30, Fig 1-13. ©Cold Spring Harbor Laboratory Press.）

在溶原周期，P_{RM}开放，而P_R和P_L关闭。抑制子协同结合到O_{R1}和O_{R2}阻碍RNA聚合酶结合P_R，抑制从该启动子开始的转录。单抑制子的结合激活了从P_{RM}开始的转录。

我们暂时回到噬菌体如何选择可变的途径这一问题上来。但首先我们要考虑诱导在细胞受到胁迫时，上面所述的溶原状态是如何转入裂解状态的。

溶原性诱导需要λ抑制子的蛋白水解切割

*E. coli*感受到DNA损害并做出反应。它通过激活RecA蛋白质的功能做到这一点。该酶参与重组（如其名称所反应的；第10章），但它也有另一功能，即刺激某一蛋

白质的自动水解切割。最先发现的这一活性的底物是一种细菌抑制子蛋白称为 LexA，它抑制编码 DNA 修复的酶。被激活的 RecA 刺激 LexA 的自我切割，解除对基因的抑制。这被称为 SOS 反应（第 9 章）。

如果细胞是溶原体，原噬菌体就倾向于在胁迫环境下逃离细胞。最终，λ 抑制子变得类似于 LexA，这使得 λ 抑制子也在 RecA 的作用下进行自身切割。切割反应除去了抑制子的 C 端结构域，因此二聚化和协同性马上不复存在。因为这两项功能对于抑制子结合 O_{R1} 和 O_{R2}（以在溶原性细胞中所见的浓度）是至关重要的，失去协同性使抑制子在这些位点（包括 O_{L1} 和 O_{L2}）上解耦联；失去抑制从而促进了从 P_R 和 P_L 开始的转录，导致了裂解生长。

为了有效诱导，溶原性细胞中的抑制子水平必须严密调控。如果水平降得太低，在正常情况下，溶原性细胞也有可能自发诱导；如果水平升得太高，诱导就会失去作用。后者的原因在于，必须有更多的抑制子失活（经由 RecA），以使其浓度降到足够低而释放 O_{R1} 和 O_{R2}。我们已经看到了抑制子是如何保证其水平不致降得太低的：它激活其自身的表达，这是一例正自我调控（positive autoregulation）。但是它是如何保证其水平不致升得太高的呢？抑制子也对自身进行负自我调控。

负自我调控（negative autoregulation）有如下作用：如图 16-26 所示，P_{RM} 被抑制子（在 O_{R2} 处）激活合成更多的抑制子。但如果浓度过高时，抑制子也会结合 O_{R3}，从而抑制 P_{RM}（以类似于裂解生长中 Cro 结合 O_{R3} 而抑制 P_{RM} 的方式）。这样阻止了新的抑制子的合成，直到其浓度降到释放 O_{R3} 的水平。

另外有趣的一点是，“诱导”一词被用来描述 λ 中由溶原性向裂解生长的转变，同时也用来描述 *lac* 基因在乳糖存在时的打开。这一通用的用词方法源自于这样一个事实，即这两个现象是同时由 Jacob 和 Monod（框 16-2）研究的。这一用法的确名实相符：乳糖诱导 Lac 的构象改变从而解除了对 *lac* 基因的抑制。同样，λ 的诱导信号也是通过造成 λ 抑制子的结构变化（此种情况下是蛋白质水解切割）而起作用的。

抑制子的负自我调控需远程相互作用和大的 DNA 环

我们已经讨论过抑制子二聚体在临近的操作子，如 O_{R1} 和 O_{R2} 上的协同结合。在溶原性细菌的原噬菌体中还有另一种协同结合。这种协同结合对于正确的负自我调控至关重要。O_{R1} 和 O_{R2} 上结合的抑制子二聚体与协同结合到 O_{L1} 和 O_{L2} 上的抑制子二聚体相互作用，形成一个抑制子八聚体；其中的每一个二聚体都独立地结合操作子。

为利于 O_R 和 O_L 之间的抑制子的相互作用，这些操作子区之间的 DNA——约 3.5kb，包括 *cI* 基因本身，必须形成一个环（图 16-27）。当环形成时，O_{R3} 和 O_{L3} 靠近，允许另外两个二聚体抑制子协同结合到这两个位点。这一协同性意味着与非协同情况下相比较，O_{R3} 结合抑制子时所需的抑制子浓度更低——实际上，只需比结合 O_{R1} 和 O_{R2} 高一点就可以了。因此，抑制子浓度是被严密控制的——很小的下降就会被其基因表达的升高所补偿，升高时，则导致基因关闭。这样就解释了为什么溶原性细菌是如此稳定而同时诱导又是非常有效的。

λ 抑制子的 C 端结构域的结构揭示了二聚体形成的基础，如早期的遗传学研究所解

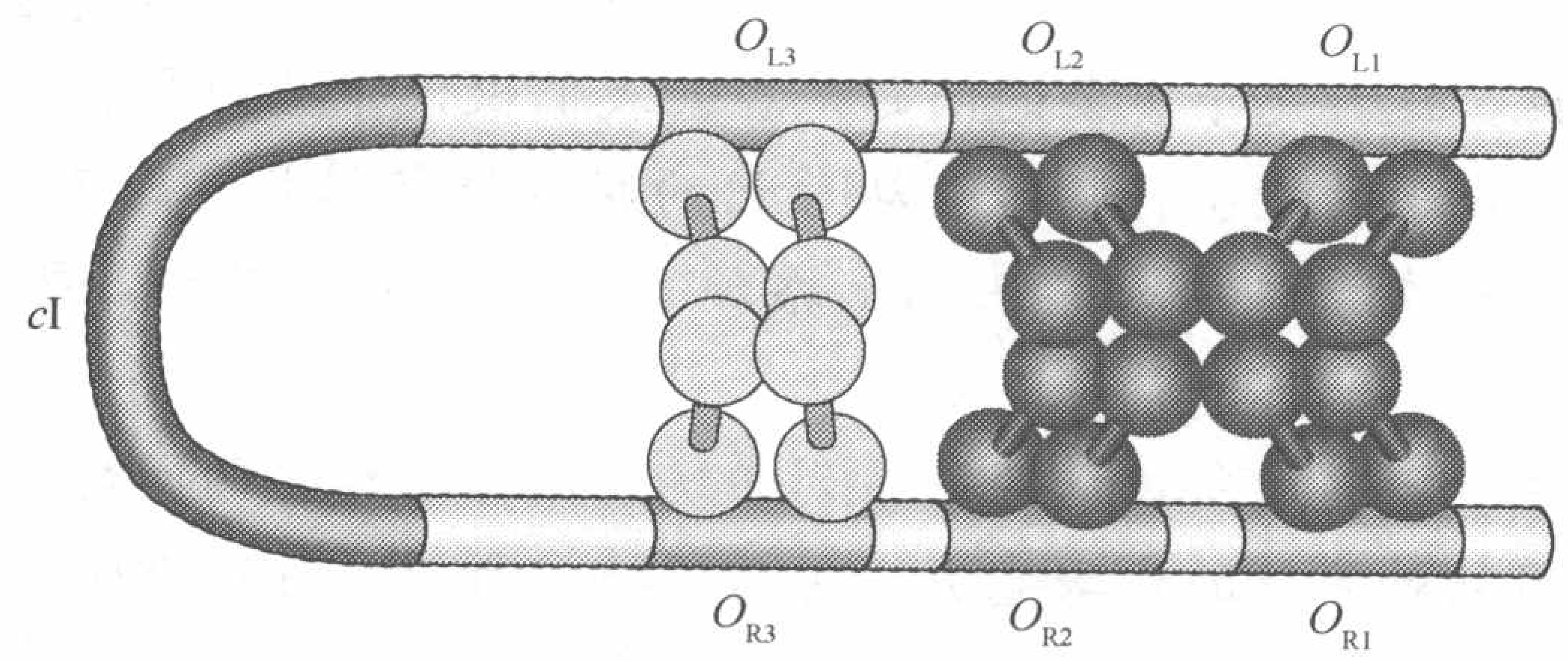


图 16-27 O_R 和 O_L 上抑制子的相互作用。这些相互作用稳定结合，这样，相互作用增强了对 P_R 和 P_L 的抑制，允许抑制子在低浓度时结合 O_{R3}。（来源：Adapted from Ptashne M. and Gann A. 2002. Genes & signals，p. 35，Fig. 1-16. ©Cold Spring Harbor Laboratory Press.）

释的。它同时也揭示了两个二聚体是如何相互作用形成四聚体的（如抑制子协同结合 O_{R1} 和 O_{R2} 时）。而且，这一结构还揭示了八聚体形成的基础——显示这是抑制子寡聚体的最高形式（图 16-28）。

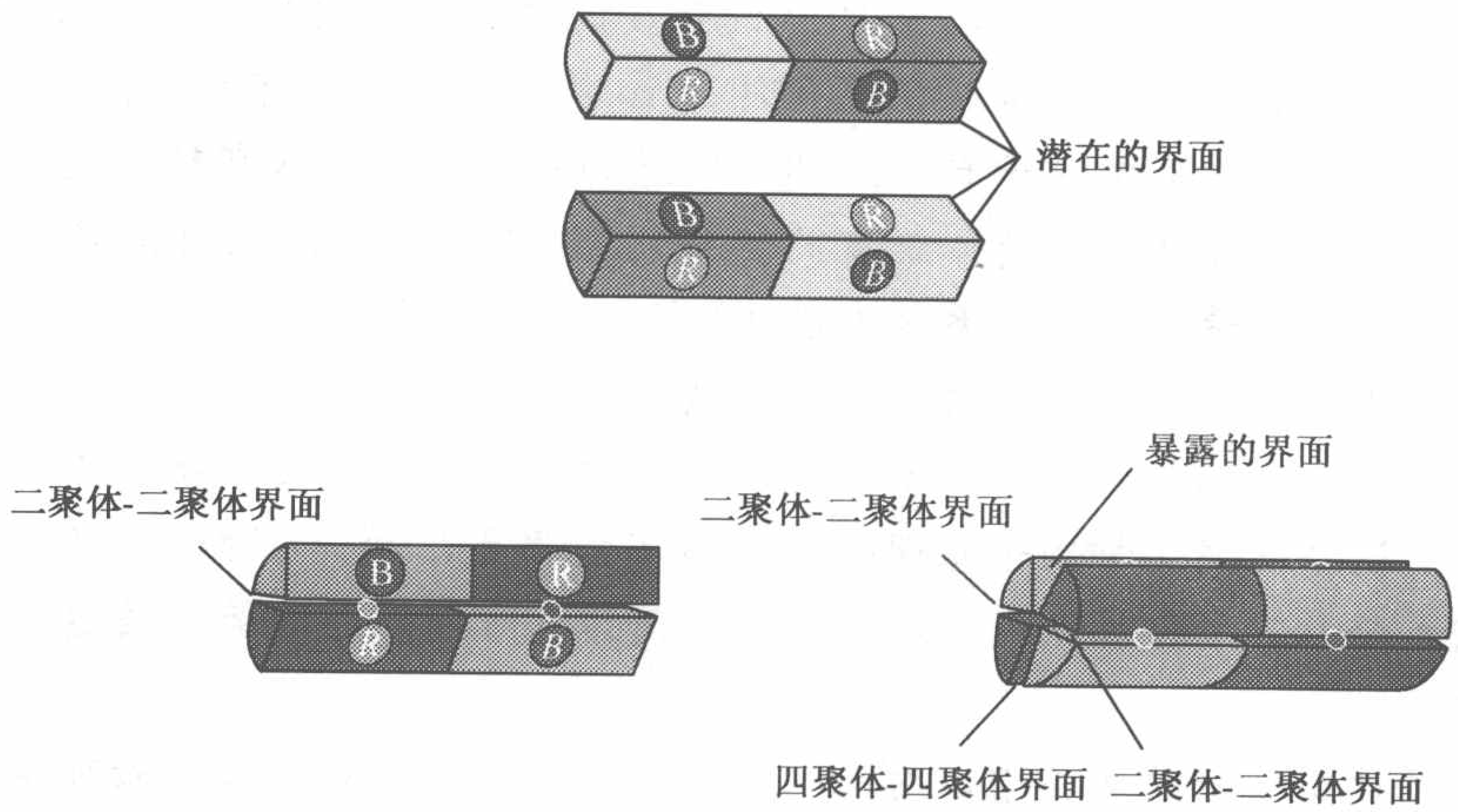


图 16-28 λ 抑制子的 C 末端结构域的互作。图的上部展示的是 2 个 λ 抑制子的 C 末端结构域的二聚体示意图。这里表示的是 λ 抑制子结构域表面形成的被称为 B 和 R 的沟，它们首先介导 2 个二聚体形成四聚体，然后 2 个四聚体又形成八聚体（当抑制子共同结合到 4 个位点 0r1，0r2，011 和 012 时可以发现这种形式。）一旦形成八聚体，就没有剩余的空间让新的二聚体进入复合体，所以八聚体是所能形成的最高形式。（经允许，略有修改，引自 Bell 等 2000，Cell 101：801-811，图 4a，b 和 5a-c Elsevier）

框 16-4 λ 转换的进化

我们重点讨论了复杂的 λ 噬菌体各种过程决定的机理：如何选择溶菌或溶原，如何有效的从稳定的原噬菌体状态转换为复制病毒的溶菌状态。形成这一特定系统的面临的许多困难已经被讨论了：如协同结合，自我正/负调控，用抑制子达到平衡的效果等。强调这些特性间的相互作用，在我们所看到的噬菌体中是如何相互依赖的，会使我们产

生这样的问题，如何从早期的原始的版本只通过很简单的步骤就形成这样的系统。对所有的生物系统来说，这都是个重要的问题，在此我们仅仅介绍的是 λ。

提出的一个分步骤的模型，揭示 λ 从原始的版本进行转换都涉及什么（框 16-4 图 1）。第一个步骤只是在系统原来的基础上添加了一个简单的内容，就可使系统工作的更好。

在最近的几年中，通过所设计的一系列的实验回答了相关的问题。这些研究旨在回答进化过程如何用相对简单的步骤完成了 λ 的转换。所以，每一个明显的特征都被突变敲掉，使噬菌体的行为发生缺陷，例如，突变的噬菌体可能溶原的效率变低或形成的溶原不稳定，或者太稳定，使侵染变得太容易或困难。

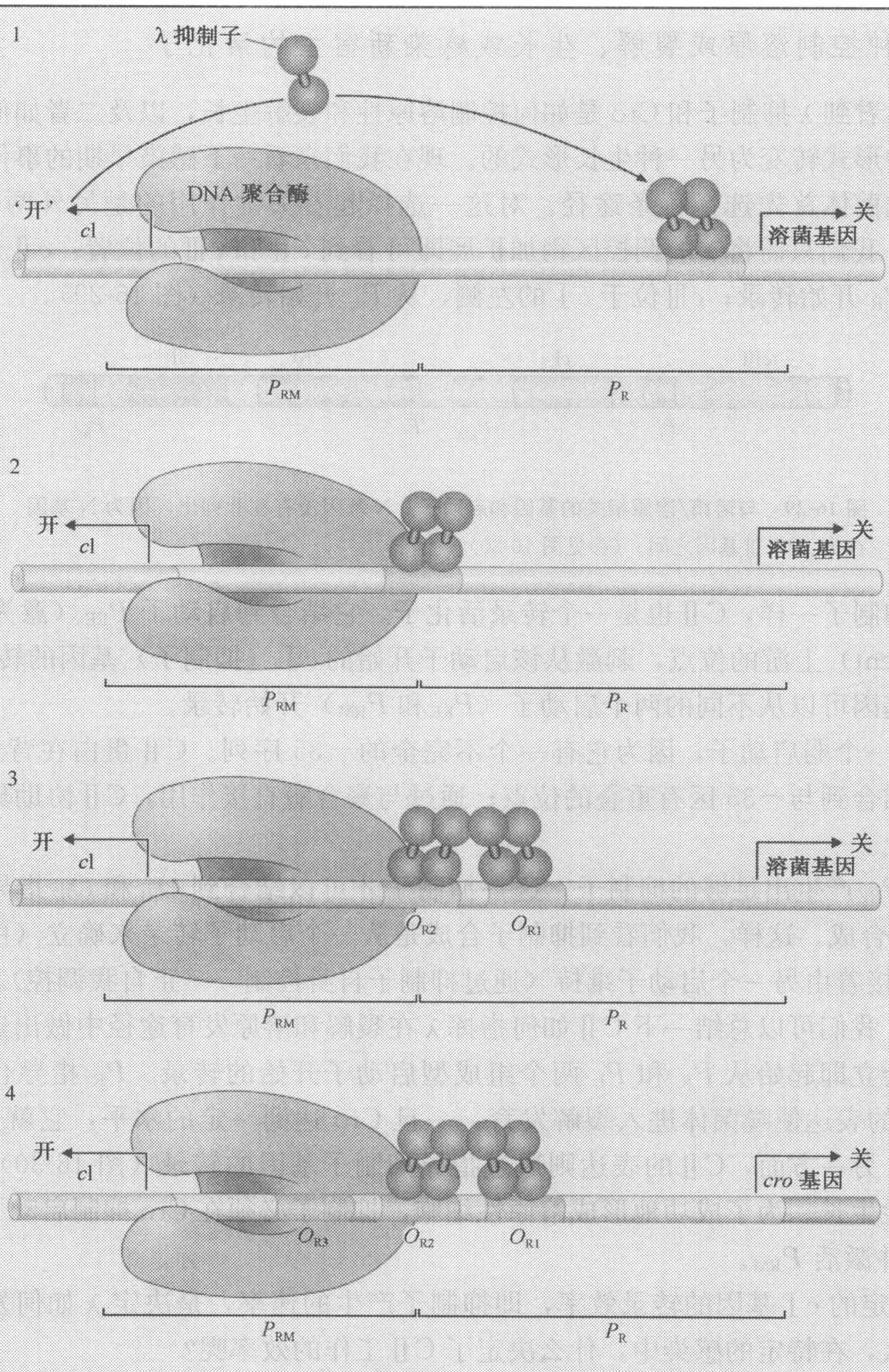
又找到了一些其它的突变可以分别补偿原来的缺陷。这些实验揭示了这个系统复杂性并非不能减少，实际上任何“重要的”特征都可被另外性状的改变所补偿，或至少部分的补偿。例如，向噬菌体的 cl 基因引入 pc 突变就会去掉正向自我调控功能。一个 pc 突变就可敲除激活子的激活功能，就向先前对 CAP 的讨论一样。所以，在这种情况下，突变的噬菌体仍然会制造抑制子，它们仍然能结合 DNA 并能抑制转录，但不能从 Prm 激活更多抑制子的转录。这种突变噬菌体能形成溶原但它们很不稳定，因为抑制子的水平很低。引入其他的变化加强启动子 Prm 的强度能够很大程度的补偿，使溶菌更稳定，更像野生型噬菌体 λ 所产生的那样。强化的 Prm 能驱使更多没有被现存抑制子激活的抑制子的表达。似乎具有抑制子自我正调控功能的野生型噬菌体有某种优势（这就解释了为什么所有已知的 λ 类的噬菌体都有此特征），但是，这对系统的正常工作不是很重要。所以，人们可以看到当前的系统在进化过程中的步骤。

还有一个例子，即抑制子的协同结合，虽然也有优点，但对噬菌体的基本生长来说也不是完全需要。所以，引入一个干扰抑制子二聚体结合的突变，就可以显著的弱化协同结合。带有这样的突变抑制子基因的噬菌体不能形成溶菌。但加入其他的修饰-加强 Prm 及其它加强 0r2 结合位点——所产生的噬菌体就可以形成溶菌，虽然比野生型 λ 的效率低一些。

在将来预计要进行的实验中，对 λ 的转换进行拆分和重新组装实验可以验证它们是如何发挥作用的以及它们是如何产生的。最近进行的有意思的实验有用细菌的抑制子蛋白质替代抑制子基因，如 Tet 抑制子，用同一噬菌体中 Cro 基因被 lacl，也就是编码 Lac 抑制子的基因替换，除此之外，将噬菌体中操作子位点进行修饰，使得两个细菌抑制子也能模仿野生型 λ 中抑制子与 Cro 的方式结合。

从这些不同片段组合起来的噬菌体也能获得野生噬菌体 λ 的部分表型。因为在修饰的噬菌体中二个抑制子的结合可由二个小分子（Lac 和 Tet 诱导子）精细的调控实现，可以通过进一步的微小的操作调查 λ 噬菌体系统是如何工作甚至如何起源的。

总而言之，通过这些不同的实验体系可以得到 2 点结论。第一，λ 转换涉及一系列的步骤，每一步在原有基础上增加一个调控的水平，使原来虽然工作，但效果略差的步骤得到改善。对任何一个系统来讲，这就是期望的自然选择结果。第二，要实现特定的目标都有变通的路线。理解这些方案最终呈现的细节比理解如何实现这些方案更困难，因为最终的方案只是众多可行方案中的一条路线。



框 16-4 图 1 推测的 λ 转换的进化阶段（阶段 1）初始的 λ 基因组有 2 个启动子，一个负责溶菌基因（P_R），一个负责抑制基因（P_{RM}）。一个与 P_R 基因重叠的 λ 抑制的结合位点，当与此位点结合的时候，抑制子关闭溶菌基因，但自身的合成并不受调控。（阶段 2）单一的抑制结合位点移动到离 P_{RM}较近的位置（即现在 O_{R2}的位置），这样就使结合的抑制子在 P_{RM}处与聚合酶相接触，从而刺激启动子抑制 P_R。（阶段 3）第二个抑制子结合位点被引入（在 O_{R1}的位置）。除此之外，还引入了一个新的蛋白质-蛋白质相互作用表面，使抑制子二聚体协同结合到这些相邻的位点。这些特征增加了系统的协同性，提高了转换的效率。（阶段 4）第三个抑制子结合位点（O_{R3}）被引入了。当此位点被结合时，抑制子负调控自已的合成，使其浓度低于关键的水平，从而保证有效的转换机制。（经允许节选自 Ptashne M. and Gann A. 1988. Curr. Biol. 8：R812-822. Elsever）

λcⅡ，一种控制溶原或裂解、生长或感染新宿主的活化子

我们已看到 λ 抑制子和 Cro 是如何控制溶原性和裂解生长，以及二者如何在受到诱导时从一种形式转变为另一种生长形式的。现在我们来看一下感染早期的事件。这些事件决定着噬菌体首先选择哪条途径。对这一选择起着关键作用的是另外两个 λ 基因，*c*Ⅱ和 *c*Ⅲ。我们只需将 λ 的调控区稍加扩展即可看到 *c*Ⅱ和 *c*Ⅲ的位置：*c*Ⅱ位于 *c*Ⅰ的右侧，从 P_R 开始转录；*c*Ⅲ位于 *c*Ⅰ的左侧，从 P_L 开始转录（图 16-29）。

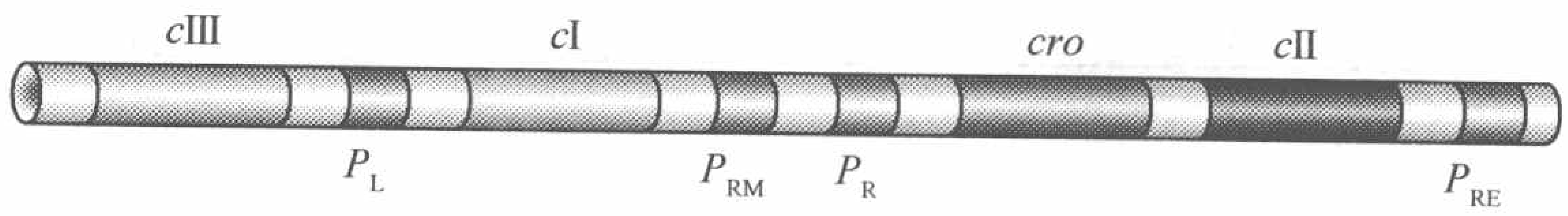


图 16-29 与溶菌/溶原相关的基因和启动子。 N 基因没有在此列出，因为 N 基因在 P_L 和 cⅢ基因之间。（参见图 16-20）

如 λ 抑制子一样，CⅡ也是一个转录活化子。它结合到启动子 P_{RE}（意为 repressor establishment）上游的位点，刺激从该启动子开始的 *c*Ⅰ（抑制子）基因的转录。因此，该抑制子基因可以从不同的两个启动子（P_{RE}和 P_{RM}）开始转录。

P_{RE}是一个弱启动子，因为它有一个不完全的－35 序列。CⅡ蛋白在背对 DNA 螺旋的一面结合到与－35 区有重叠的位点；通过与聚合酶直接作用，CⅡ协助聚合酶结合到启动子上。

只有 P_{RE}产生出足够的抑制子，这一抑制子才可以结合到 O_{R1}和 O_{R2} 指导其自身从 P_{RM}开始的合成。这样，我们看到抑制子合成是从一个启动子转录来确立（由一个活化子激活），接着由另一个启动子维持（通过抑制子自身控制——正自我调控）。

现在，我们可以总结一下 CⅡ如何指挥 λ 在裂解和溶原发育途径中做出选择。一经感染，就会立即起始从 P_R 和 P_L 两个组成型启动子开始的转录。P_R 指导合成 Cro 和 CⅡ。Cro 的表达使噬菌体进入裂解发育：一旦 Cro 达到一定的水平，它就会结合 O_{R3} 封闭 P_{RM}。另一方面，CⅡ的表达则通过指导抑制子基因的转录（图 16-30）使噬菌体进入溶原性生长。为了成功地形成溶原性细胞，抑制子必须在 Cro 抑制启动子之前结合 O_{R1}和 O_{R2}并激活 P_{RM}。

CⅡ决定的 cⅠ基因的转录效率，即抑制子产生的速率，是决定 λ 如何发展的关键步骤。那么，在特定的感染中，什么决定了 CⅡ工作的效率呢?

就像框 16-5（鉴别与溶菌/溶原相关基因的遗传途径）所描述的过程，精心设计的遗传筛选获得了这些 cⅠ，cⅡ和 cⅢ基因。

侵染特定细菌的噬菌体颗粒的数目影响侵染走向溶菌或溶原

感染复数（或称 moi）是度量有多少噬菌体颗粒侵染一个群体中的特定细菌的单位。如果平均每个细胞的噬菌体颗粒数目是 1 或几，侵染很可能以溶菌结束。如果噬菌体颗粒的数目是 2 个或多个，则很可能形成溶原。当每个细胞中的噬菌体颗粒数目减少时，形成溶菌的可能性就变大，当数目不断增加时，形成溶原的可能性增加。

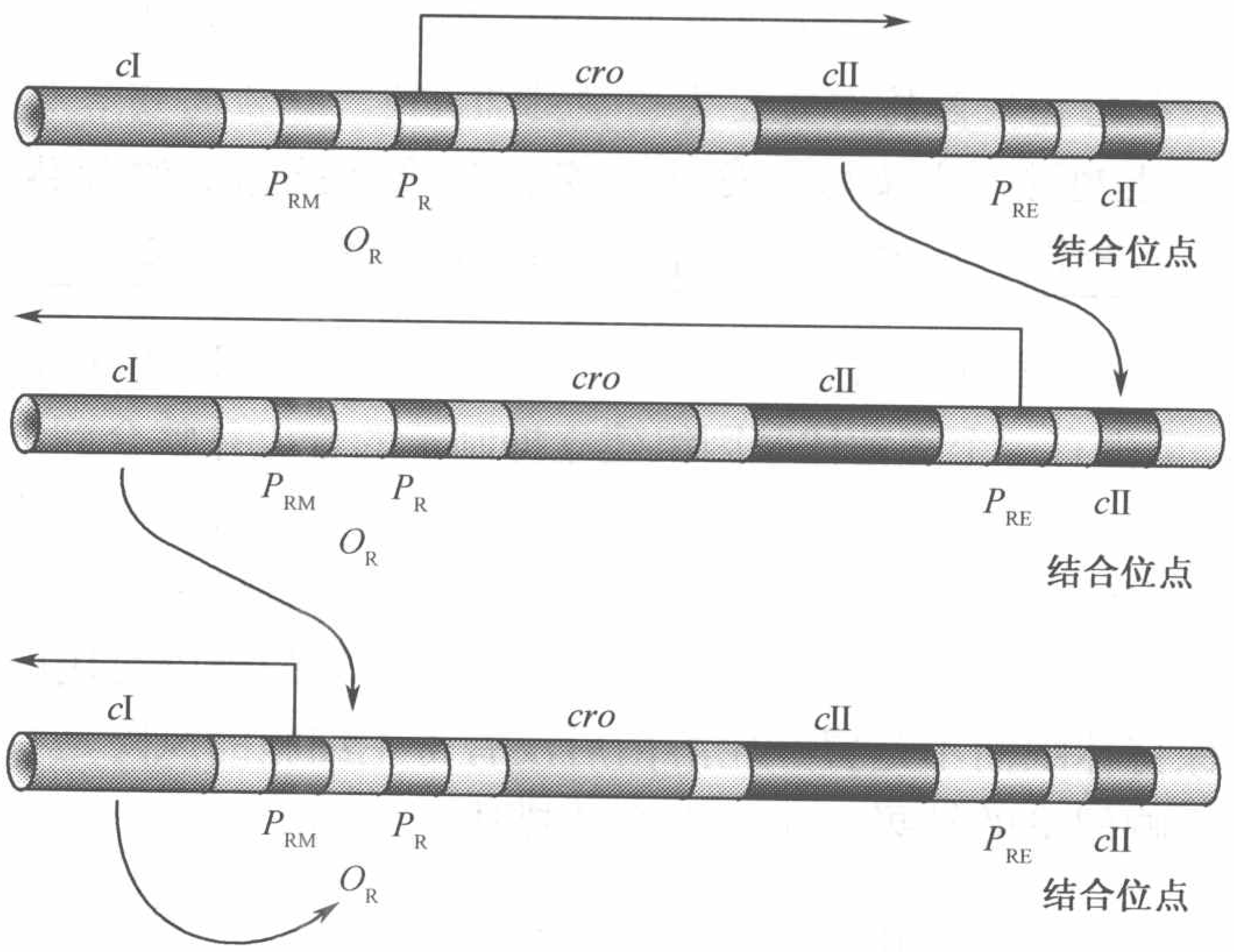


图 16-30 溶原性的建立。当建立了溶原性时，*c*Ⅰ基因从 P_{RE}开始转录，当维持住这一状态时，*c*Ⅰ基因从 P_{RM}开始转录。结合在 O_{R1}和 O_{R2}上的抑制子不但激活了维持模式的表达而且关闭了建立模式的表达。注意 P_R 不但控制了裂解基因的表达，它还控制了 *c*Ⅱ的表达，因此它对于溶原和裂解生长都是重要的。类似地，尽管图中没有表示出来，P_L 控制了许多裂解基因也同时控制了协助建立溶原性的 *c*Ⅲ基因（见正文）。（来源：Adapted from Ptashne M. and Gann A. 2002. Genes & signals, p. 31, Fig. 1-14. ©Cold Spring Harbor Laboratory Press.）

框 16-5 遗传学方法确定参与裂解/溶原选择的基因

通过筛选分别只能有效地进行裂解或溶原生长的λ突变体确定了参与裂解/溶原选择的基因。为了理解是如何发现这些突变体的，我们要考虑噬菌体是如何在实验室中生长的（第 21 章）。细菌细胞可以在琼脂培养板上长成一个汇合在一起、不透明的菌落。长在菌落上的裂解噬菌体产生透明斑或洞（图 21-3）。每个斑一般是由一个噬菌体感染细菌细胞形成的。这次感染产生的后代细胞感染周围细胞，如此这般，杀死（裂解）最初受到感染的细胞附近的细菌细胞并在不透明的菌落中造成一个透明的无细胞区。

噬菌体λ也形成噬菌斑，但这些斑是模糊的（或雾状的），即斑的内部比未受感染的菌落更透明一些，但只在边上未受感染。其原因在于，λ不像纯粹的裂解噬菌体，它们只杀死一部分的感染细胞，其他的作为溶原体存活下来。溶原体对于接下来的感染有抗性，因此可以不被噬菌斑中的大量噬菌体颗粒伤害而生长。这种“免疫力”的原因很简单：在溶原体中，整合的噬菌体 DNA（原噬菌体）不断从 P_{RM}合成抑制子。任何进入细胞的λ基因组都立即被抑制子结合，没有机会进行裂解生长。

在一次经典的研究中，分离了形成透明斑的λ突变体。这些突变噬菌体不能形成溶原体但仍能进行裂解生长。λ透明突变体确认了 3 个噬菌体基因，称为 *c*Ⅰ、*c*Ⅱ、*c*Ⅲ（意为 clear Ⅰ、Ⅱ、Ⅲ）。其他研究分离出所谓的恶性（*vir*）突变体。在这些突变体中确认了λ抑制子结合的操作子位点，有鉴于这样的噬菌体可以溶原体生长这一事实分离得到。通过与 *lac* 系统比较，发现 cⅠ突变体类似于 Lac 抑制子（*lac*Ⅰ）突变体，*vir* 突变

体与 *lac* 操作子（*lacO*）突变体相同（框 16-2）。另一不同的实验中还发现了又一启发性的突变，这是一个宿主基因的突变。突变被称为 *hfl*，意为高频溶原（high frequency lysogeny）。当被野生型的 λ 感染时，这一突变系几乎总是形成溶原体，很少允许噬菌体裂解生长。这一细菌株系缺少降解 λ *c*Ⅱ蛋白的蛋白酶（见正文）。

这是有道理的。进入细胞的噬菌体基因组越多，并开始从 Pr 和 pl 起始转录，制备的 CⅡ和 CⅢ就越多，至少有一个噬菌体基因组建立抑制子的合成并整合到细菌基因组中的几率就越大。只要有一个侵染的噬菌体做到了这一点，其他的就相应的被阻止发生溶菌。

人们可以推测为什么 λ 要如此表现，例如，为什么当有很多噬菌体但只有很少细菌时它宁可发生溶原。如果只有很少的细菌细胞，供下一轮的侵染可用的寄主细胞受到限制，所以噬菌体保持休眠而不是冒一轮溶菌后没有寄主细胞的风险是有好处的。细菌细胞生长的状态也影响侵染的结果，正如下文的描述。

E. coli 的生长条件控制 CⅡ 的稳定性并因而控制裂解/溶原的选择

当噬菌体感染的是一群健康的、生长快速的细菌细胞时，它倾向于通过裂解繁殖，将后代释放到富含宿主细胞的环境中。而当条件对细菌生长不利时，噬菌体就更有可能形成溶原细胞并保持原有状态。附近只有很少的宿主细胞可以被后代噬菌体感染，这些不同的生长条件以如下方式影响 CⅡ。

CⅡ在 *E. coli* 中是一个非常稳定的蛋白质，它被一个特异的蛋白酶 FtsH（HflB）降解，这个酶由 *hfl* 基因编码。因此，CⅡ指导合成抑制子的速度就由它被 FtsH 降解的速度来决定。缺少 *hfl* 基因的细胞（也就是缺少 FtsH）几乎总是在被 λ 感染时形成溶原细胞：在无蛋白酶时，CⅡ是稳定的，它能指导合成足够的抑制子。FtsH 自身的活性由细菌细胞的生长条件调控，尽管我们不知道这是如何实现的，但我们至少知道如下内容：当生长良好时，FtsH 活性高，CⅡ被有效破坏，抑制子不能合成，噬菌体倾向于裂解生长；在恶劣条件下则相反，FtsH 活性低，CⅡ降解慢，抑制子积累，噬菌体倾向于溶原生长。CⅡ的水平也被噬菌体蛋白 CⅢ调节。CⅢ稳定 CⅡ，这可能是因为它可以作为 FtsH 的替代性（因此是竞争性）底物。

第二个 CⅡ蛋白依赖的启动子 P_I，有一类似 P_{RE} 的序列，位于噬菌体基因 *int*（图 16-20）之前。这一基因编码一种酶，该酶催化位点特异性重组，使 λ DNA 整合入细菌染色体形成原噬菌体（第 11 章）。第三个 CⅡ依赖的启动子 P_{AQ}，位于基因 *Q* 的中间，其作用是延滞裂解发育从而促进溶原发育。这是因为 P_{AQ} 的 RNA 可以作为翻译信使结合到 Q 信使促进其降解。Q 是另一个调控蛋白，它促进晚期裂解生长，我们将在以下部分介绍。

λ 发育中的转录抗终止作用

前面我们已经介绍了在转录起始后的步骤中运作基因调控的例子。在 λ 发育中还发现了另外两种例子，下面我们将进行描述。首先，我们将看一种正转录调控的类型，叫做抗终止作用（antitermination）。

在无N蛋白和Q蛋白时，由这两个调控蛋白控制的转录也可以很好地起始。但是，除非调控蛋白修饰RNA聚合酶，否则在启动子下游几百到一千核苷酸处转录就会终止。λ、N和Q蛋白因此被称为抗终止蛋白。

N蛋白调控早期基因表达，它作用于3个终止子：一个在*N*基因本身的左侧，一个在基因*cro*右侧，还有一个在基因*P*和*Q*之间（图16-20和图16-31）。Q蛋白有一个靶标，是一个位于晚期启动子$P_{R'}$200核苷酸下游的终止子，位于*Q*和*S*基因之间（图16-31）。λ晚期基因操纵子，自$P_{R'}$开始转录，在原核转录单位中算的是很大的，大约26kb，RNA聚合酶要走完全程需要10min。

我们对抗终止蛋白的作用机制的了解是不彻底的。如其他调控蛋白一样，N和Q也只作用于带有特定序列的基因。N蛋白阻止在λ早期操纵子的终止，但不作用于其他细菌和噬菌体操纵子。没有在抗终止蛋白作用的位点发现抗终止蛋白识别的特异序列，但在终止子之前确实发生了抗终止作用。N蛋白所需的位点叫做*nut*（意为N利用，*N* utilization)，位于P_L和P_R下游60～200核苷酸处（图16-31）。但是N不在DNA内结合这些序列，而是结合从包含有*nut*序列的DNA转录出来的RNA。

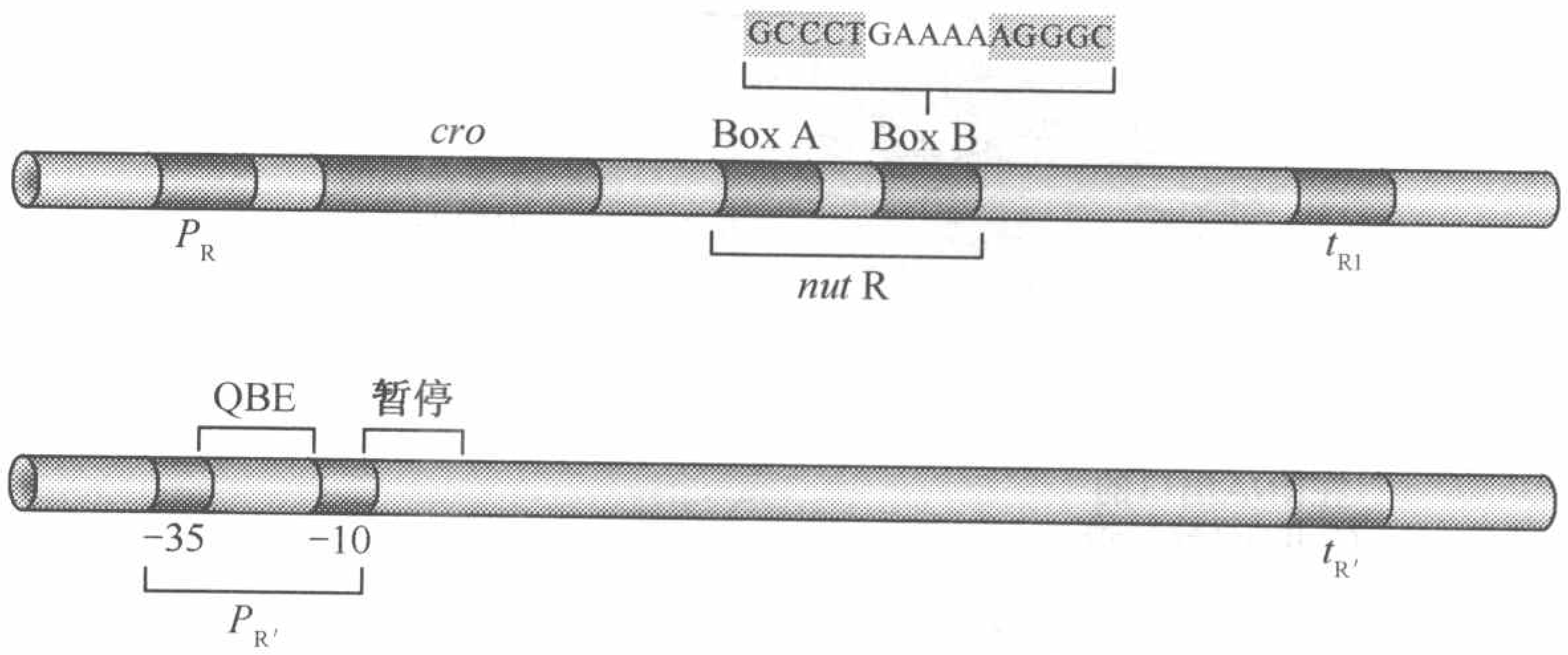


图16-31 λ转录抗终止子N和Q的识别位点和激活位点。上一行显示早期的右向启动子P_R及其起始终止子，t_{R1} *nut*位点分为两个区，称为BoxA（7bp）和BoxB，由一个8bp的间隔区域分开。BoxB的序列成对称并在转录为RNA后形成一个茎-环结构。RNA样的*nut*R的序列如上面所示。下面一行显示启动子$P_{R'}$，其序列为Q蛋白功能所必需，此外还有Q蛋白作用的终止子。

因此，RNA聚合酶一通过*nut*位点，N即结合到RNA上，并于那里装载到RNA聚合酶上。在这种状态下，聚合酶就能够抵抗*N*和*cro*基因以外的终止子。λN与细菌基因*nusA*、*nusB*、*nusE*和*nusG*的产物一起作用。NusA蛋白是一个重要的细胞转录因子；NusE是核糖体小亚基蛋白S10，但它在N蛋白功能中的作用还是未知的；NusB蛋白在细胞中的功能也不清楚。这些蛋白质与N蛋白在*nut*位点形成一个复合体，但是如果N蛋白浓度高，也可以在没有这些蛋白质的时候起作用，表明N蛋白自身具有促进抗终止作用。

不像N蛋白，λQ蛋白识别位于晚期启动子$P_{R'}$－10和－35区之间的DNA序列(QBE)（图16-31）。在无Q时，聚合酶结合并起始转录，但在行进仅16或17核苷酸后暂停；接着它会继续直到终止于下游大约200bp的终止子（$t_{R'}$）。如果存在Q，聚合

酶一离开启动子，Q 就会结合 QBE，并转到在附近暂停的聚合酶。聚合酶载有了 Q 就能通过 $t_{R'}$ 继续转录。

最近，聚合酶的 σ 因子与 Q 功能的相关得到了证实（参见图 16-32）。首先，聚合酶在 $P_{R'}$ 处刚刚起始就停顿下来是因为它遇到了与启动子－10 区相似的一段序列。σ 因子的第二区域通常识别这一序列，如 12 章所描述的那样与非模板链的碱基对结合。在停顿区也发生相同的事情，使聚合酶暂时停止。与此同时，新生的转录本穿过酶的 RNA 通道使 σ 和酶的核心交界处发生重排，暴露原来被掩盖的部分 σ 因子的区域 4（如 12 章的描述）。σ 的这个表面然后与 Q 结合。

现在仍然不清楚为什么聚合酶与 Q 形成的新的复合体不能越过下游的终止子。但有一点是清楚的，就是 σ 从开始就可以调控下游基因，并且 σ 的区域 4 自始至终都可以是调控子的靶点。

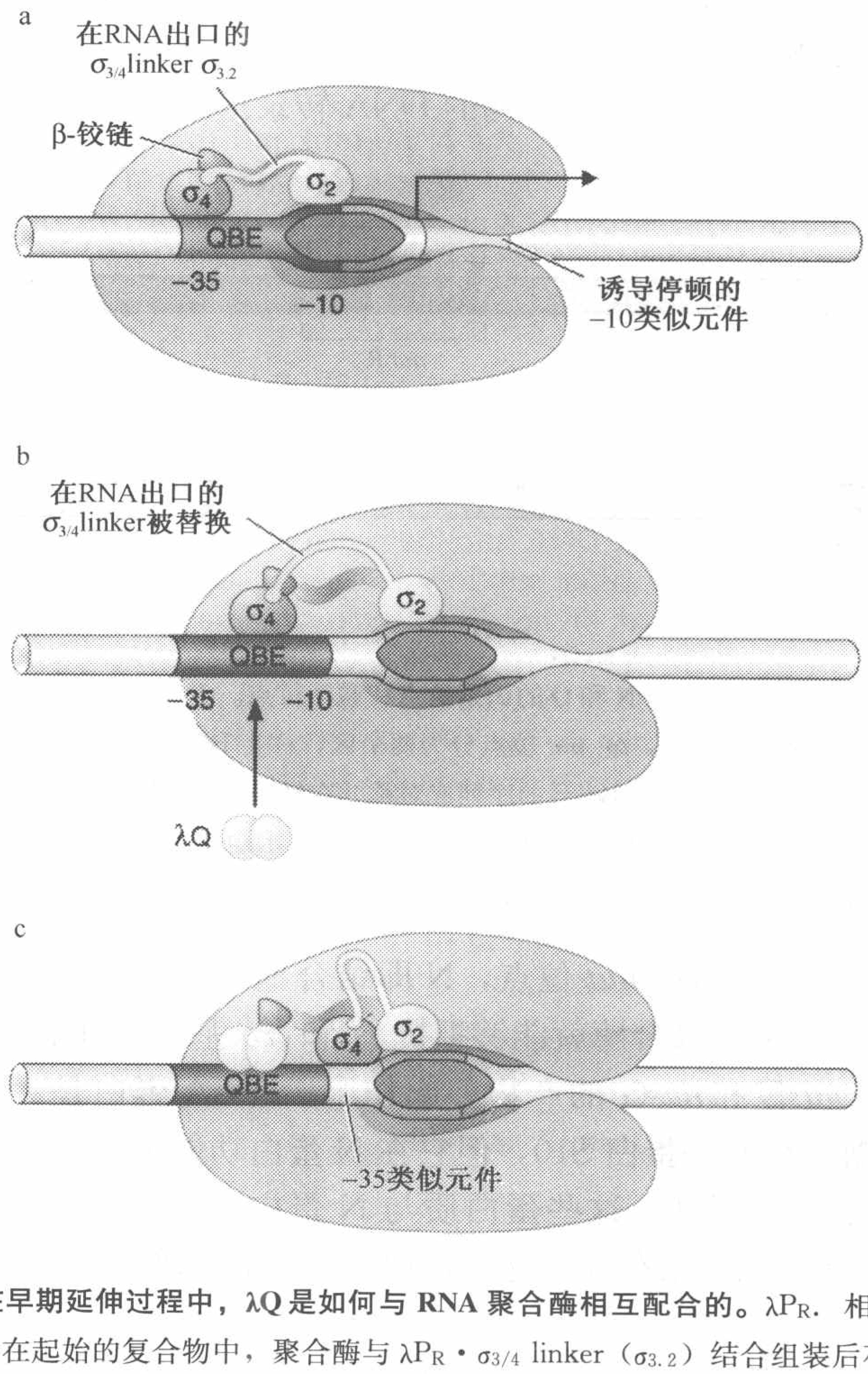


图 16-32 在早期延伸过程中，λQ 是如何与 RNA 聚合酶相互配合的。$λP_R$. 相关事件的发生顺序（a）在起始的复合物中，聚合酶与 $λP_R$ · $σ_{3/4}$ linker（$σ_{3.2}$）结合组装后存在于 RNA 出口中。（b）停顿的复合物。新生的转录本（未展示）替换了出口通道中的 $σ_{3/4}$ linker，这个事件发生在与 λQ 结合之前。（c）λQ 表现为与停顿的延伸因子结合。细节描述详见正文。（Ann Hochschild 提供）

逆回调控：RNA合成和稳定性控制相互影响并决定基因表达

CⅡ蛋白激活启动子 P_I，指导 *int* 基因的表达，同时激活启动子 P_{RE}，指导合成抑制子（图 16-20）。Int 蛋白是一个在形成溶原性细胞时将噬菌体基因组整合到宿主细胞中的酶（第 11 章）。因此，一旦感染，有利于 CⅡ蛋白活性的条件就造成抑制子和整合酶的猛增。

但是 *int* 基因可以从 P_L 或 P_I 转录，所以或许有人会想到整合酶也可以在无 CⅡ蛋白的时候合成。这种情况是不会发生的，原因在于由 P_L 起始的 *int* 信使 RNA 被细胞的核酸酶降解了，而在 P_I 起始的 mRNA 是稳定的，可以被翻译为整合酶蛋白质。这是因为两种信使的 3′端有着不同的结构。

在 P_I 起始，并在 *int* 基因末端之后 300 核苷酸处的终止子停止的 RNA，具有典型的茎-环结构，其后是 6 个尿嘧啶（图 16-33；第 12 章图 12-12）。相反，当 RNA 的合成从 P_L 起始，同时 RNA 聚合酶被 N 蛋白修饰，此时通过终止子。这一更长的 mRNA 可以形成一个茎，作为核酸酶的底物。因为这一负责负调控的位点位于它所影响的基因的下游，同时降解是沿着基因逆向进行的，所以这一过程被称为逆向调控（retroregulation）。

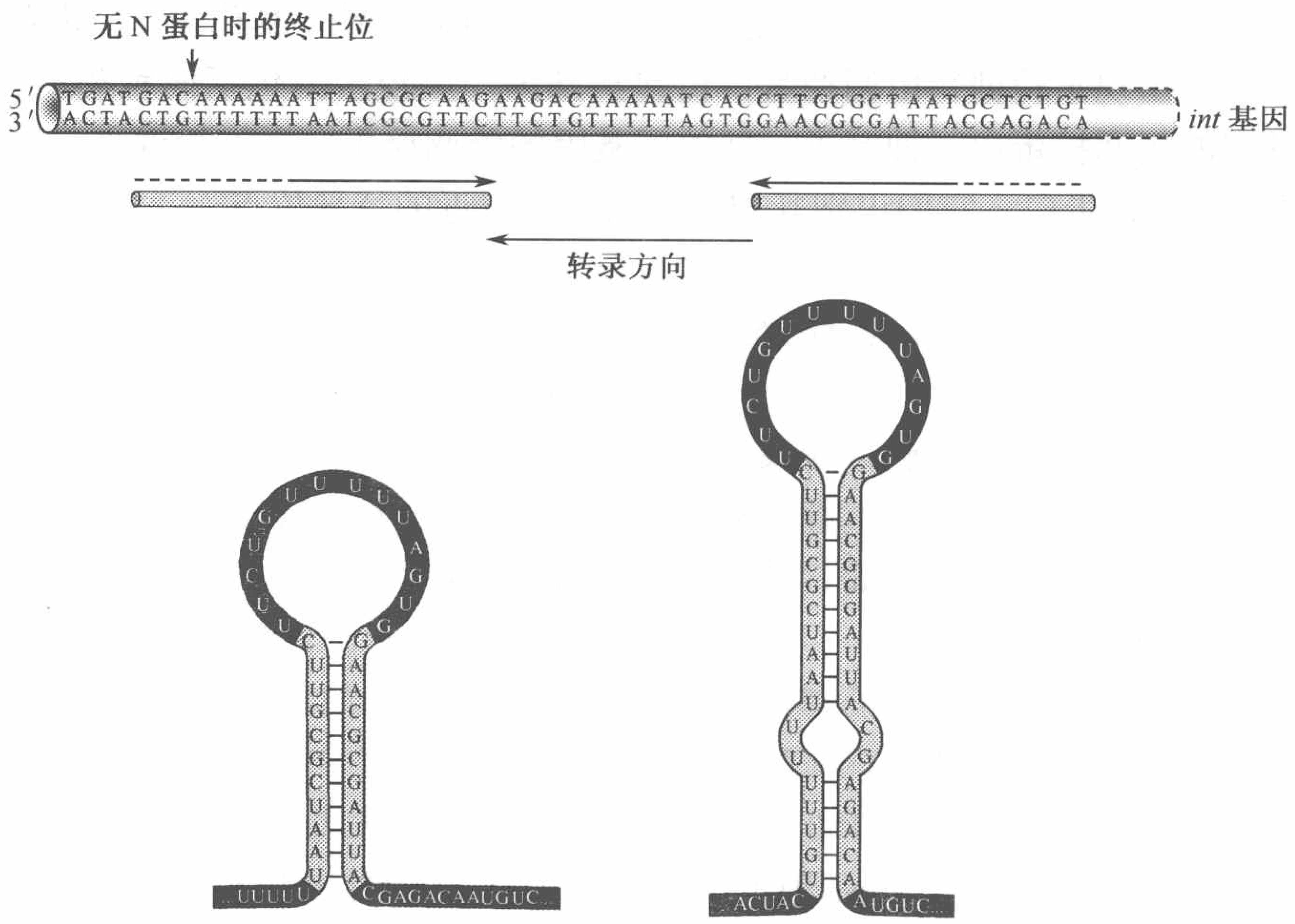


图 16-33 *int* 表达的逆向调控的 DNA 位点和转录 RNA 的功能结构。上部所示为 DNA 序列，下面的圆柱体显示 RNA 中形成发夹对称的序列。左侧的结构显示从 $P_{I'}$ 转录出的 RNA 形成的终止子，未被 N 蛋白拮抗终止，这一结构可以抵抗核酸酶的降解。右侧的结构显示在 N 蛋白抗终止因子影响下从 P_L 转录出的 RNA 形成的延伸环，这一结构显示 RNA 酶Ⅲ切割和核酸酶降解的靶位。

逆向调控的生物学功能是不言而喻的。当 CⅡ活性低而有利于裂解发育时，是不需要整合酶的。因此，其 mRNA 被破坏掉。但当 CⅡ活性高而有利于溶原发育时，*int* 基

因表达促进受抑制的噬菌体 DNA 重组进入细菌染色体。

这一调控机制还有另一层次的意义。当原噬菌体被诱导时，它需要合成整合酶和另一个酶，称为剪切酶（excisionase），见第 11 章，通过重组，催化噬菌体 DNA 从细菌 DNA 中重新独立出来。不管 CⅡ活性高不高，它都要做到这一点。因此，在这些条件下，不管 N 蛋白的活性如何，噬菌体必须从 P_L 合成稳定的整合酶 mRNA。这是如何实现的呢？

当噬菌体基因组在溶原型细胞建立期间整合到细菌染色体时，发生重组的噬菌体黏附位点位于 *int* 基因和编码 mRNA 开始降解的延伸茎的序列之间（图 16-20）。因此，在整合形式下，引起降解的位点在 *int* 基因末端被删除，从 P_L 转录出来的 *int* mRNA 是稳定的。

小结

一般的基因根据对其产物的需要开启和关闭。这一调控主要发生在转录起始水平。例如，在 *E. coli* 中，编码代谢乳糖的酶的基因只有在培养基中有乳糖时，才高水平地转录。而且，当有葡萄糖（一种更好的能源）时，即使存在乳糖，其代谢基因也不表达。

信号（如出现某种糖）由调控蛋白传递给基因。调控蛋白包括两种类型：活化子，即开启基因的正调控蛋白；抑制子，即关闭基因的负调控蛋白。一般说来，这些调控蛋白都是 DNA 结合蛋白，它们识别其控制基因上或附近的特异位点。

活化子，在最简单（也是最常见）的情况下，作用于弱启动子。这些启动子，在无任何调控蛋白存在时，RNA 聚合酶结合上去（并因此启动转录）的效率比较低。活化子以其一面结合到 DNA，另一面结合到聚合酶并将其募集到启动子部位。这一过程是一个协同结合的例子，足以刺激转录。

抑制子通过结合到启动子重叠区从而阻碍 RNA 聚合酶的结合，可以抑制转录。抑制子也可以其他方式作用。例如，结合启动子旁的位点，通过与结合到启动子上的聚合酶相互作用抑制转录。

E. coli 的 *lac* 基因被活化子和抑制子以前面描述过的简单方式控制。在无葡萄糖时，CAP 结合 *lac* 启动子附近的 DNA，通过募集 RNA 聚合酶到启动子，激活这些基因的表达。Lac 抑制子结合与启动子有重叠区的位点，在无乳糖时关闭基因表达。

RNA 聚合酶被募集到不同基因上的另一种方式是利用 σ 因子。不同的 σ 因子可以替代主要的 σ 因子（*E. coli* 中是 σ^{70}）并将聚合酶引导到启动子不同的序列上。例如，σ^{32} 指导对热激反应的基因的转录，σ^{54} 指导参与氮代谢酶基因的转录。噬菌体 SPO1 利用一系列的选择性 σ 控制其基因在感染期间的有序表达。

在细菌中还有其他种类的转录激活的例子。在有些启动子上，RNA 聚合酶独立、有效的结合，并形成稳定的、无活性的封闭复合体。这一封闭复合体不会自动地转变为开放复合体并起始转录。在这样的启动子上，活化子必须刺激从封闭复合体到开放复合体的转变。

刺激此类启动子的活化子通过变构起作用：它们与稳定的封闭复合体相互作用，诱导构象变化，导致转变为开放复合体。在本章，我们看到了两个活化子通过变构激活转录的例子。其中一例，活化子（NtrC）与RNA聚合酶（带有σ^{54}）结合到*glnA*启动子的稳定封闭复合体上，刺激转变为开放复合体。在另一例中，活化子（MerR）诱导*merT*启动子中的DNA构象改变。

在我们考虑过的所有情况中，调控蛋白本身都被信号通过变构控制，即调控蛋白的形状在信号存在时改变。一种状态下，它可以结合DNA，另一种状态则不能。例如，Lac抑制子就被配体别乳糖（从乳糖而来的产物）控制。当别乳糖结合了抑制子，它就诱导蛋白产生形状变化，使之处于不能结合DNA的状态。

基因表达也可在转录起始后的步骤中调控。例如，可在转录延伸水平调控。在此，我们讨论了3种情况：*trp*基因的衰减作用以及λ噬菌体N蛋白和Q蛋白的抗终止作用。*trp*基因编码合成色氨酸所需的酶。这些基因只在缺少色氨酸时才被转录。一种氨基酸控制这些基因表达的方式是衰减作用。转录在*trp*启动子起始，如果细胞中存在色氨酸（以Trp tRNA的形式），并在结构基因之前终止，那么λ噬菌体的N蛋白和Q蛋白就装载到RNA聚合酶上在噬菌体基因组某启动子起始转录。RNA聚合酶以这种方式被修饰之后就可以通过某些转录终止位点。否则这些转录终止位点会阻止对其下游基因的表达。除了转录之外，我们还看到了一例基因调控是怎样作用于mRNA翻译水平的（我们讨论的实例是核糖体蛋白基因）。

在本章的结论部分我们详细讨论了噬菌体λ是如何在两种选择性的繁殖方式之间进行选择的。在这一系统中遇到的基因调控的方式也在其他系统中作用，包括如我们将在以下章节中看到的。例如，动物发育控制中，协同结合就被用于严格控制开关效应，以及建立和维持基因表达的不同途径。

（宋其峰　译　刘国振　刘斌　校）

参考文献

书

Baumberg S., ed. 1999. *Prokaryotic gene expression.* Oxford University Press, Oxford.

Müller-Hill B. 1996. *The* lac *operon.* de Gruyter, Berlin.

Ptashne M. 2005. *A genetic switch: Phage lambda revisited,* 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Ptashne M. and Gann A. 2002. *Genes & signals.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

激活与抑制

Adhya S., Geanacopoulos M., Lewis D.E., Roy S., and Aki T. 1998. Transcription regulation by repressosome and by RNA polymerase contact. *Cold Spring Harbor Symp. Quant. Biol.* **63:** 1–9.

Busby S. and Ebright R.H. 1999. Transcription activation by catabolite activator protein (CAP). *J. Mol. Biol.* **293:** 199–213.

Dodd I.B., Shearwin K.E, and Egan J.B. 2005. Revisited gene regulation in bacteriophage λ. *Curr. Opin. Genet. Dev. Biol.* **15:** 145–152.

Dove S.L., Darst, S.E., and Hochschild A. 2003. Region 4 of σ as a target for transcription regulation. *Mol. Microbiol.* **48:** 863–874.

Gottesmann M. and Wesiberg R. 2004. Little lambda, who made thee? *Microbiol. Mol. Biol. Rev.* **68:** 796–813.

Hochschild A. and Dove S.L. 1998. Protein–protein contacts that activate and repress prokaryotic transcription. *Cell* **92:** 597–600.

Huffman J.L. and Brennan R.G. 2002. Prokaryotic transcription regulators: More than just the helix-turn-helix motif. *Curr. Opin. Struct. Biol.* **12:** 98–106.

Jacob F. and Monod J. 1961. Genetic regulatory mechanisms

in the synthesis of proteins. *J. Mol. Biol.* **3:** 318–356.

Lawson C.L., Swigon D., Murakami K.S., Darst S.A., Berman H.M., and Ebright R.H. 2004. Catabolite activator protein: DNA binding and transcription activation. *Curr. Opin. Struct. Biol.* **14:** 10–20.

Magasanik B. 2000. Global regulation of gene expression. *Proc. Natl. Acad. Sci.* **97:** 14044–14045.

Müller-Hill B. 1998. Some repressors of bacterial transcription. *Curr. Opin. Microbiol.* **1:** 145–151.

Murray N.E. and Gann A. 2007. What has phage lambda ever done for us? *Curr. Biol.* **17:** R305–R312.

Oppenheim A.B., Oren Kobiler O., Stavans J., Court D.L., and Adhya S. 2005. Switches in bacteriophage lambda development. *Ann. Rev. Genet.* **39:** 409–429.

Ptashne M. 2006. Lambda's switch: Lesson from a module swap. *Curr. Biol.* **16:** R459–R462.

Ptashne M. and Gann A. 1997. Transcriptional activation by recruitment. *Nature* **386:** 569–577.

Rojo F. 2001. Mechanisms of transcriptional repression. *Curr. Opin. Microbiol.* **4:** 145–151.

Rombel I., North A., Hwang I., Wyman C., and Kustu S. 1998. The bacterial enchancer-binding protein NtrC as molecular machine. *Cold Spring Harbor Symp. Quant. Biol.* **63:** 157–166.

Roy S., Garges S., and Adhya S. 1998. Activation and repression of transcription by differential contact: Two sides of a coin. *J. Biol. Chem.* **273:** 14059–14062.

Schleif R. 2003. AraC protein: A love–hate relationship. *Bioessays* **25:** 274–282.

Wigneshweraraj S.R., Burrows P.C., Bordes P., Schumacher J., Rappas M., Finn R.D., Cannon W.V., Zhang X., and Buck M. 2005. The second paradigm for activation of transcription. *Prog. Nucleic Acid Res. Mol. Biol.* **79:** 339–369.

Xu H. and Hoover T.R. 2001. Transcriptional regulation at a distance in bacteria. *Curr. Opin. Microbiol.* **4:** 138–144.

DNA结合、协同性及变构效应

Bell C.E. and Lewis M. 2001. The Lac repressor: A second generation of structural and functional studies. *Curr. Opin. Struct. Biol.* **11:** 19–25.

Hochschild A. 2002. The switch: *c*I closes the gap in autoregulation. *Curr. Biol.* **12:** R87–R89.

Luscombe N.M., Austin S.E., Berman H.M., and Thornton J.M. 2000. An overview of the structures of protein–DNA complexes. *Genome Biol.* **1:** REVIEWS001.

Monod J. 1966. From enzymatic adaptation to allosteric transitions. *Science* **154:** 475–483.

Vilar J.M.G. and Saiz L. 2005. DNA looping in gene regulation: From the assembly of macromolecular complexes to the control of transcriptional noise. *Curr. Opin. Genet. Dev. Biol.* **15:** 136–144.

抗终止

Artsimovitch I. 2005. Control of transcription termination and antitermination. In *The bacterial chromosome* (ed. N.P. Higgins), Chapter 17. ASM Press, Washington, D.C.

Friedman D.I. and Court D.L. 2001. Bacteriophage λ: Alive and well and still doing its thing. *Curr. Opin. Microbiol.* **4:** 201–207.

Gottesman M. 1999. Bacteriophage λ: The untold story. *J. Mol. Biol.* **293:** 177–180.

Greenblatt J., Mah T.F., Legault P., Mogridge J., Li. J., and Kay L.E. 1998. Structure and mechanism in transcriptional antitermination by the bacteriophage λ N protein. *Cold Spring Harbor Symp. Quant. Biol.* **63:** 327–336.

Roberts J.W., Yarnell W., Bartlett E., Guo J., Marr M., Ko D.C., Sun H., and Roberts C.W. 1998. Antitermination by bacteriophage λ Q protein. *Cold Spring Harbor Symp. Quant. Biol.* **63:** 319–325.

Weisberg R.A. and Gottesman M.E. 1999. Processive antitermination. *J. Bacteriol.* **181:** 359–367.

第17章 真核生物的基因调控

本章概要

•

从酵母到哺乳动物转录调控的保守机制

•

通过真核生物活化子募集结合基因的蛋白复合物

•

信号整合与组合控制

•

转录抑制子

•

信号转导与转录调控子的控制

•

组蛋白与 DNA 修饰导致的基因“沉默”

•

基因的表观遗传调控

我们讲述过的在细菌中基因表达调节的方式在真核细胞中都能见到，此外真核细胞还具有一些细菌中所不具备的基因表达调节方式。在这些方式中，最显著的应为剪接（splicing）。如我们在第 13 章所见，很多情况中一个特定的转录本可以通过不同的剪接方式产生多种剪接产物，而且这也是可调控的。

同细菌一样，调控步骤最常见于转录起始阶段。涉及生物体内被调控的转录过程中的许多原理确实同样适用于真核生物的转录调控。这些原理在原核生物基因调节一章（第 16 章）的开始部分已有叙述，并且在该章的末尾已经总结。我们热切地希望那些先前（或最近）没有阅读过这些内容的读者，在继续本章学习之前能够先阅读这些章节。

我们也已注意到，真核生物比细菌拥有更加精细的转录系统（第 12 章）。这在 RNA 聚合酶Ⅱ系统（用于转录蛋白编码基因）中尤为如此。尽管复杂度增加，转录还受活化子（activator）和抑制子（repressor）的调节（它们是一类 DNA 结合蛋白，能够在特定基因上对适当信号做出反应，来促进或阻止转录的起始）。然而，真核细胞和基因的额外特征使得这些调控蛋白的作用变得复杂。我们将通过总结这些额外特征中最重要的两个特征来开始本章的学习。

核小体及其修饰物：正如我们在第 7 章所提到的，真核生物的基因组包裹在组蛋白中，形成核小体。因此，转录调控是在一种反应底物部分被包裹的情况下进行的。这一情形使得在没有转录调控蛋白存在的情况下，许多基因的表达量减少。真核细胞也包含了许多能对组蛋白进行重排或化学修饰的酶，这种修饰改变核小体，使转录机器及 DNA 结合蛋白易于结合底物并进行转录。因此，核小体引出了一个在细菌中没有遇到

的问题，但对核小体的修饰也为调控提供了新的机会。

更多的调控子和更密集的调控序列：真核细胞与原核细胞间更进一步的区别是控制特定基因的调控蛋白数目上的差异，这一点从基因调控蛋白结合位点的数目和排列上可以看出。在细菌中，单个调控蛋白与小段序列结合，而在真核细胞中，与细菌相比，这种结合位点数目更为庞大，位置更加远离转录起始点。我们把基因上转录系统结合的区域称作启动子（promoter），单一结合区域称作调控蛋白结合位点（regulator binding site），一个基因上包含所有调控蛋白结合位点的 DNA 片段称作调控序列（regulatory sequence）。

调控序列的扩展，即一个基因调控蛋白结合位点数目的增加，在果蝇、哺乳动物等多细胞生物中尤其显著。这反映了在这些生物中更加广泛的信号整合，即在合适时空条件下基因开关时要求更多信号的趋势。我们在细菌中见过信号整合的例子（第 16 章），但这些例子仅是典型的两个不同的调节蛋白通过整合两个信号来控制一个基因（如葡萄糖和乳糖对 *lac* 基因的调节）。酵母的信号整合比多细胞生物少（在这方面它们与细菌其实没有多大的差别），而且，它们基因上的调控序列也比多细胞的真核生物要少（图 17-1）。

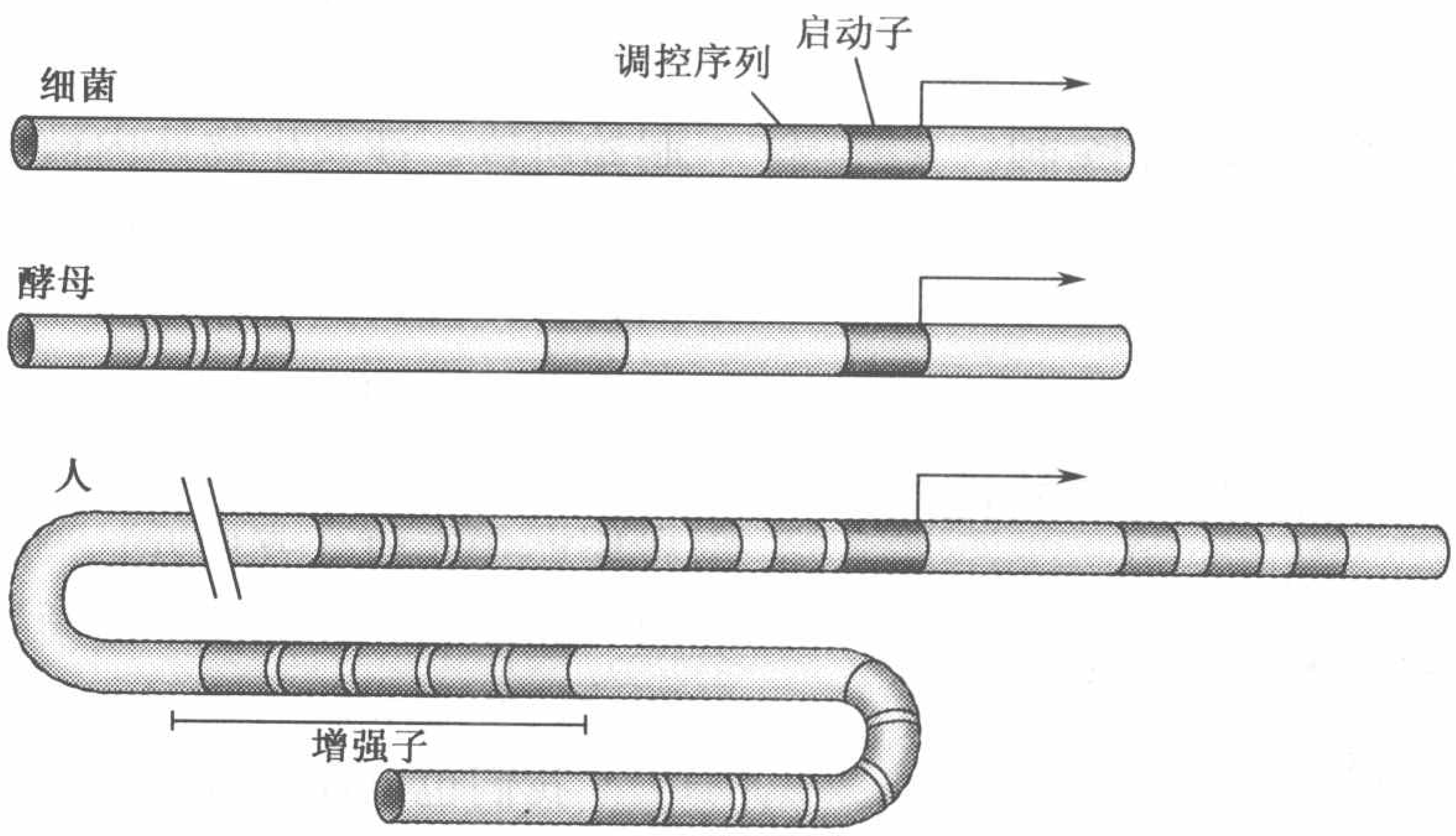


图 17-1　细菌、酵母和人类基因组的调控元件。图例说明了调控序列复杂性的增加。从受单个抑制子调控的简单细菌基因，到受多个活化子和抑制子调控的人类基因。每个例子都标出了启动子，这是转录起始的位点。它在细菌中位置确定，而在真核细胞中，转录起始位点位于转录复合物结合位点的下游（第 12 章）。在人类的调控序列中存在一类调控结合位点被称为增强子，如图所示。

在多细胞生物中，调控序列可以延伸至距启动子数千个核苷酸处（上游或下游均可），并可由数十个调控蛋白结合位点组成。通常这样一组结合位点被称作增强子（enhancer），一个增强子结合的调控蛋白将负责在一个合适时空状态激活基因表达。不同的增强子与不同组的调控蛋白结合，控制同一基因在不同时空的表达，以应答不同的信号。

拥有更广泛的调控序列意味着某些调控蛋白结合位点将远离它们所控制的基因，有

时能达到 50kb 或更远。在这么远的距离上，调控蛋白是如何作用的呢？在细菌中我们看到距离几个 kb 的 DNA 结合蛋白能进行通信：如 λ 噬菌体中，O_R 位点和 O_L 位点上的抑制子能相互作用；又如 NtrC 能从 1kb 或更远处激活 *glnA* 基因。在这些“远距作用”的例子中，居中的 DNA 回路为蛋白质之间的相互作用提供了方便。即使在某些情况下蛋白质作用的距离很远并且形成回路的原因尚不清楚，但同样的机制也可以解释真核细胞中存在的许多远距离调控的情况，虽然不能针对所有可能的情况。

远距的激活产生了另外一个问题。当结合于一增强子时，也许有几个基因都在此活化子的作用范围之内，但一个特定的增强子通常只调节一个基因。已在增强子和启动子之间找到了其他的调控序列，称为绝缘子（insulator）或边界元件（boundary element）。绝缘子阻止结合于增强子的活化子对启动子的激活。尽管目前对这些元件的了解尚不深入，但它们确保了活化子的工作不会产生混乱。

从酵母到哺乳动物转录调控的保守机制

本章我们将考虑从单细胞的酵母到哺乳动物的基因调控。所有这些生物都具有真核生物两个典型的特征，即精细的转录调控系统和核小体及其修饰物。因此，在所有的真核生物中很多基因调控的基本属性是一致的就不足为奇了。酵母是最适于同时进行遗传学和生物化学研究的生物，大量的有关活化子或抑制子的信息正是来自于该生物。另外，对于将其结论普适化最为关键的是，当典型的酵母活化子在哺乳动物细胞内表达时能够激活转录。这是通过报道基因（reporter gene）来检测的。报道基因含有酵母活化子结合位点，它插入到某一基因启动子的上游，同时此基因的表达水平易于检测（后面还将讨论）。

我们将看到典型的真核细胞活化子的工作方式与最简单的细菌非常相似：DNA 结合区域和激活区域分开，并通过募集蛋白复合物到特定的基因上来激活转录过程。相反，真核细胞中的抑制子能以多种方式工作，这与我们在细菌中所见有所不同。这些方式包括基因沉默（gene silencing），即募集核小体及 DNA 修饰物到基因组的区域来保持基因甚至大片段 DNA 的封闭。

尽管有许多共同的特征，但并不是在所有真核生物中基因调控的所有细节都是一致的。更重要的是，我们曾提到过，与多细胞生物相比，一个典型酵母基因的调控序列要短一些。所以我们必须研究更高级的生物，来了解基因调节的基本机制是如何扩展以适应更为复杂的信号整合和组合式控制的。本章将把讨论局限于蛋白质（及其修饰）介导的转录调控，下一章将讨论 RNA 分子介导的基因表达调控。

活化子的 DNA 结合与激活功能的分离

在细菌中考察一个典型的活化子，如 CAP，我们发现它的 DNA 结合与激活功能是分离的。我们这样描述这个遗传示例：阳性控制（*pc*）的突变能正常地与 DNA 结合，但不能激活基因。真核细胞活化子的 DNA 结合区域与激活区域也是分离的。事实上，

本例中，蛋白质的两个表面通常位于分离的两个功能域。

我们以最常研究的真核活化子 Gal 4 为例（图 17-2）。该蛋白质能激活酿酒酵母（*Saccharomyces cerevisiae*）半乳糖基因的转录。与细菌中的半乳糖基因类似，这些基因编码半乳糖代谢所必需的酶。其中一个基因名为 *GAL1* 。Gal4 能与 *GAL1* 上游 275bp 处的 4 个位点结合（图 17-3）。结合后，如果存在半乳糖，Gal4 将激活 *GAL1* 的转录，使转录效率提高 1000 倍。

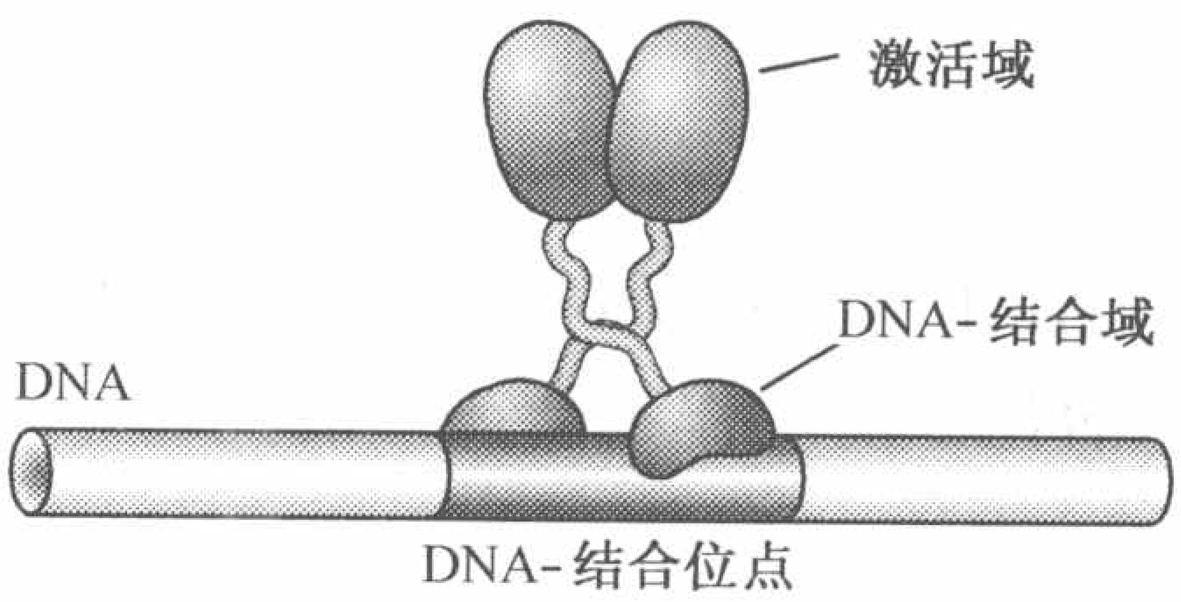


图 17-2 Gal4 与 DNA 上位点的结合。酵母活化子 Gal4 以二聚体的形式与 DNA 上一段 17bp 的位点结合。该蛋白质的 DNA 结合区域与激活区域（激活域）是分离的。

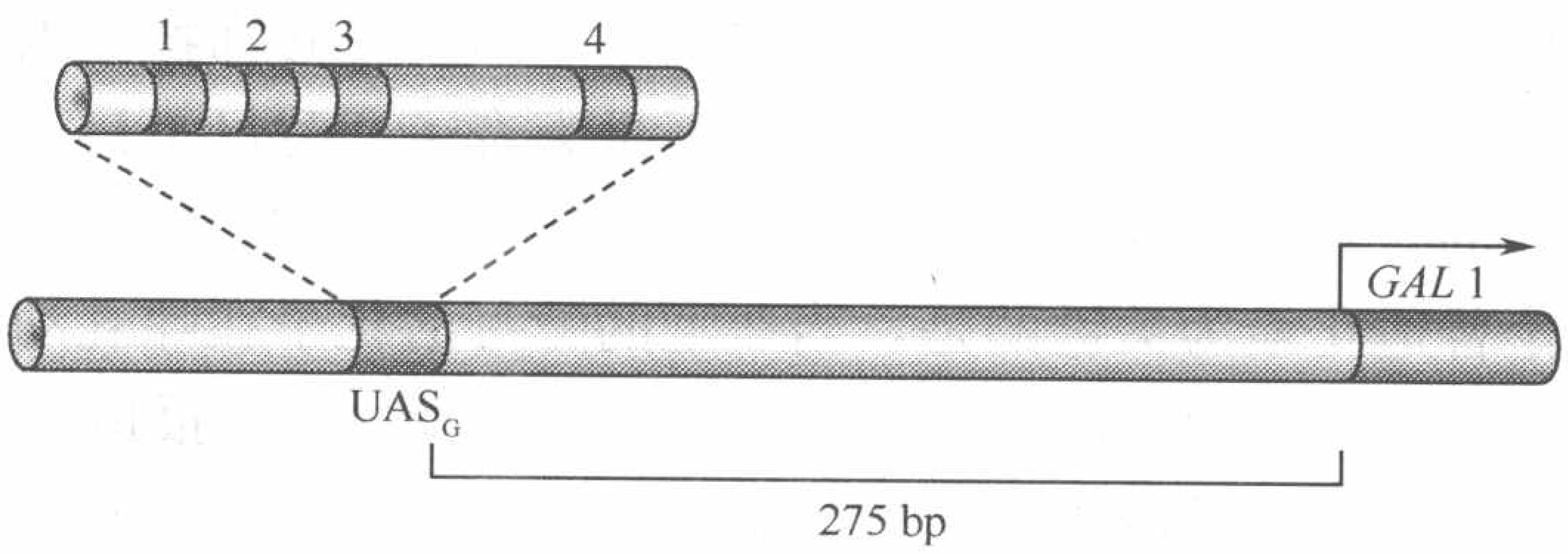


图 17-3 酵母 *GAL1* 基因的调控序列。GAL 的上游激活序列（UAS$_G$）包括 4 个位点，如图 17-2 所示，每一位点与一个 Gal4 二聚体结合。尽管没有显示，在 UAS$_G$ 和 *GAL1* 基因之间还存在着另一位点 Mig1，该位点能与抑制子结合，我们将在本章后部分进一步提到它。（图 17-21）

两个补充实验揭示了 Gal4 的 DNA 结合域与激活域的分离。在其中一个实验中，表达 *GAL4* 基因的片段（编码这个活化子 N 端的 1/3），得到的产物能与 DNA 正常结合，但是不能激活转录过程。这样的蛋白质包含 DNA 结合域，但是缺乏激活域，因此形式上相当于细菌活化子的 *pc* 突变（图 17-4a）。

另一个实验构建了一个杂合的基因，该基因将编码 Gal4 蛋白 C 端 3/4 的序列，同细菌抑制子 LexA 的 DNA 结合域融合。产生的蛋白质在酵母中同一个报道质粒一起表达，该质粒 *GAL1* 启动子的上游具有 LexA 结合位点。融合蛋白能激活该报道质粒的转录（图 17-4b）。实验表明，激活不单单是通过与 DNA 结合来介导的，如同我们在细菌中遇到的机制一样，即通过 MerR 的激活发挥作用（图 16-16）。相反，DNA 结合域的作用仅仅是将激活域与启动子联系在一起，就如同我们在细菌中遇到的最常见的机制一样。

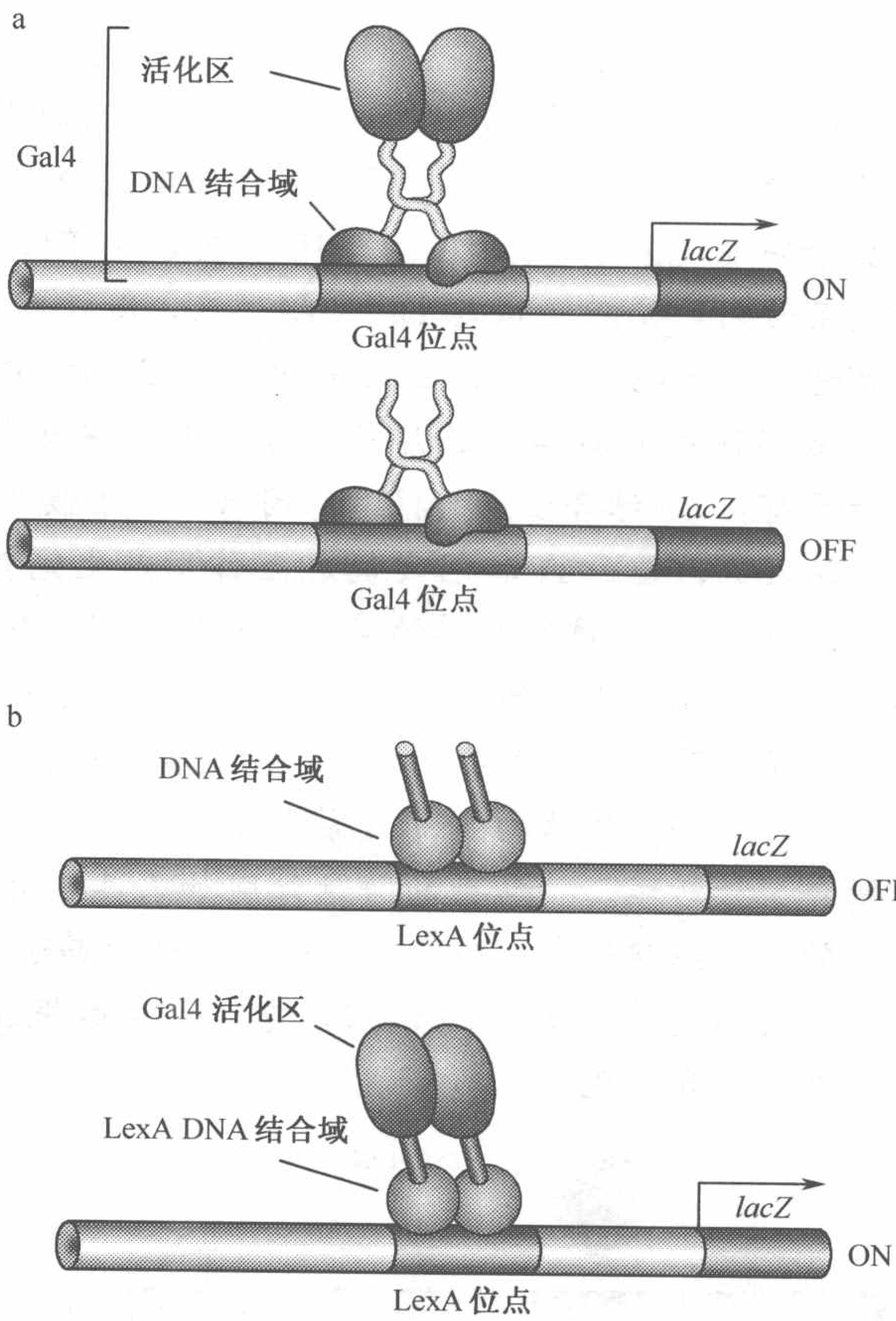
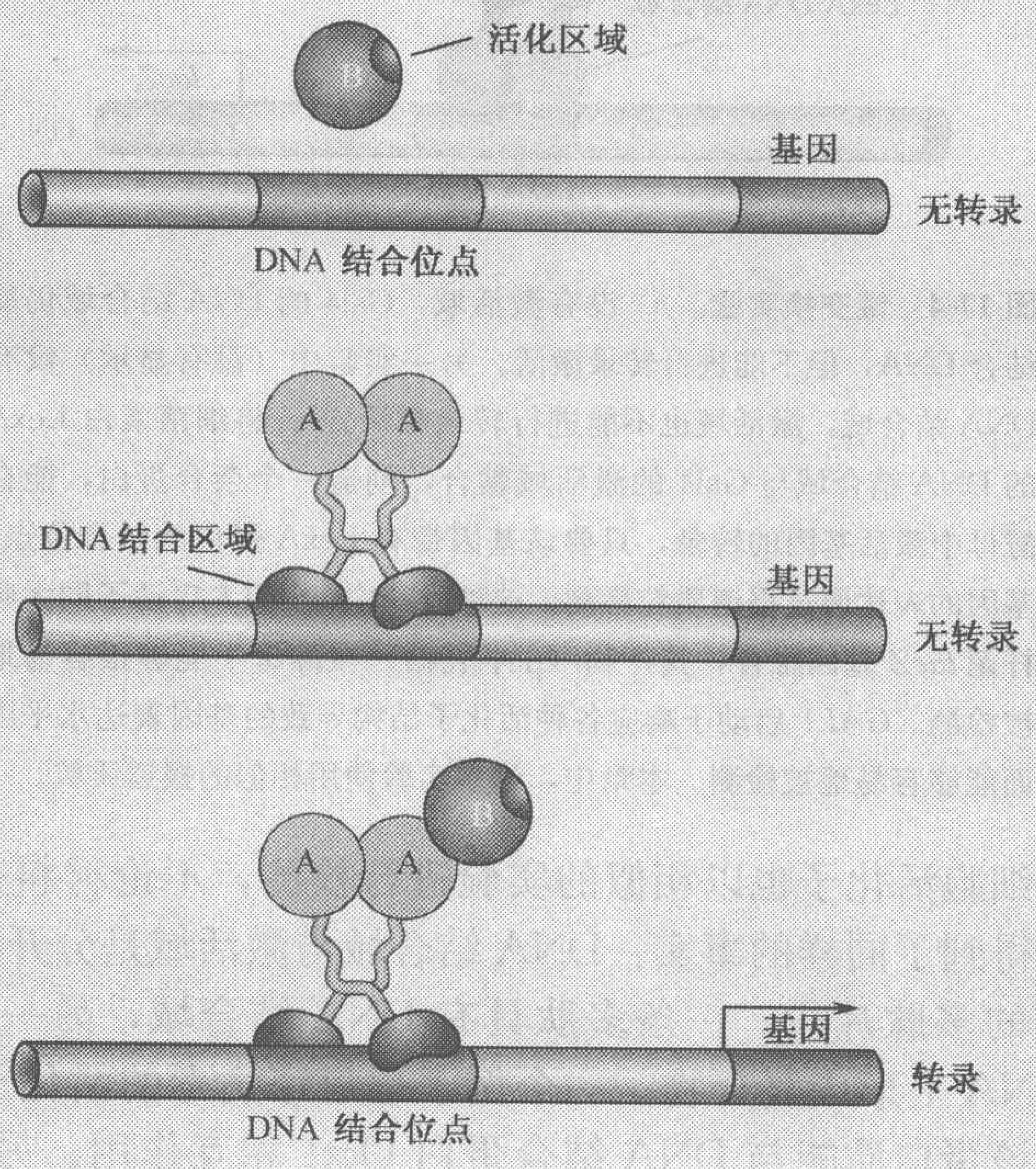


图 17-4 域交换实验。a. 没有激活域，Gal4 的 DNA 结合域仍能结合 DNA，但不能进行转录激活。另一实验中（没有显示）没有 DNA 结合域，激活域也不能进行转录激活。b. 将细菌蛋白 LexA 的 DNA 结合域与 Gal4 的激活域黏合，创造一个杂合蛋白，能在酵母中激活基因的转录，只要该基因带有 LexA 蛋白的结合位点。基因的表达通过报道质粒检测，这种质粒中，*GAL1* 启动子同大肠杆菌 *lacZ* 基因融合，其产物（β-半乳糖）在酵母中能够很容易地被检测。*GAL1* 启动子响应各种活化子结构导致的基因表达水平因而能够容易地被检测。本章中，许多实验使用相似的报道质粒。

很多其他真核细胞活化子也以相似的实验进行检测，不论材料是来自酵母、果蝇，还是哺乳动物，都得到了同样的事实：DNA 结合域与激活域是分开的。在某些例子中，它们甚至位于不同的多肽片段，一条多肽具有 DNA 结合域，另一条多肽则具有激活域，两条多肽在 DNA 上形成复合物。一个例子是疱疹病毒的活化子 VP16，在感染该病毒的细胞中发现该蛋白质能与 DNA 结合蛋白 Oct1 相互作用。另一个例子是果蝇的活化子 Notch，我们将在第 19 章中对此做介绍。真核细胞活化子的 DNA 结合域与激活域分离的特点是一个检测蛋白质间相互作用的实验能够广泛运用的基础（框 17-1，双杂交实验）。

框 17-1 双杂交实验

本实验用于鉴定蛋白质间的相互作用。在框 17-1 图 1 的示例中，激活报道基因的前提是蛋白 A 同蛋白 B 发生作用（即使这些蛋白质本身不在转录激活中起作用）。如正文中讨论的一样，本实验的发现表明，DNA 结合域与激活域可位于不同的蛋白质上，只要这些蛋白质之间相互作用，使激活域与 DNA 相联系，激活附近的基因。实际上，本实验通过以下步骤来进行。编码蛋白 A 的基因同编码 *Gal4* 的 DNA 结合域的片段融合，另一蛋白 B 的编码基因同编码激活域的片段融合。当在酵母中单独表达，任何一种融合蛋白都不能激活具有 *Gal4* 结合位点的报道基因（如图 1 最开始两行显示）。当两种融合蛋白在酵母中同时表达，蛋白 A 和 B 之间相互作用产生完整的活化子，使报道基因表达（如图 1 最后一行显示）。这个简单的实验被广泛运用于从库中筛选候选蛋白，以找到与某种已知的起始蛋白相互作用的蛋白质。图中蛋白 A 可看作为起始蛋白（“诱饵”），蛋白 B(“猎物”)即代表库中众多供候选者之一（见第 21 章关于库构建的描述）。酵母细胞转染了两种质粒，一个编码结构蛋白 A 融合了 DNA 结合域，另一个是库，编码众多未知的蛋白质融合了激活域。被转染的酵母细胞如果其 A 和 B 蛋白能相互作用将激活报道基因。这样的一种细胞形成一个克隆，这个克隆能够通过置于有合适指示剂的培养基中来鉴定。典型的报道基因是 *lacZ*，阳性克隆（表达报道基因的细胞群）将在有适当指示剂的培养皿上呈现蓝色。



框 17-1 图 1 双杂交技术工作原理。本技术中使用的报道基因典型的是 *lacZ* 或其他产生易于检测的基因。

真核细胞调控蛋白使用一系列 DNA 结合域，但 DNA 识别的原理与细菌的相同

上述实验表明，以细菌的 DNA 结合域代替真核生物活化子的 DNA 结合域，同样能够行使功能。提示 DNA 结合蛋白识别它们的结合位点的途径在不同的生物体内没有本质上的差异。

从前面的章节中我们知道大多数细菌的调控蛋白是以二聚体的形式与双螺旋对称的 DNA 目标序列结合，每个单体将一段 α 螺旋嵌入 DNA 大沟中，覆盖该位点一半以上的区域，紧靠着碱基对的边缘部分。结合过程通常不需要蛋白质和 DNA 结构变化。大多数细菌调控蛋白采用螺旋-转折-螺旋域。如我们所见，这种域由两个被一段短的转折隔开的 α 螺旋组成。一个螺旋（识别螺旋）同 DNA 大沟相契合，并识别特殊的碱基对。另一个螺旋同 DNA 骨干相联系，为识别螺旋正确定位，并增强结合力（图 5-20）。

尽管有细节上的差异，但大多数真核细胞采用同样的 DNA 识别原理，即蛋白质经常以二聚体形式与 DNA 结合，并以 α 螺旋嵌入大沟来识别特定的 DNA 序列。典型的真核调节蛋白的识别螺旋非常类似于螺旋-转折-螺旋域；其他真核调节蛋白识别螺旋的域结构相当不同。我们在原核细胞中没有见到过的变化有，几个在真核细胞中见到的调控蛋白以异二聚体（heterodimer）的形式与 DNA 结合，在某些例子中，甚至是以单体的形式与 DNA 结合（尽管经常要在其他蛋白的协同下起作用）。异二聚体增加了可使用的 DNA 结合特异性的范围：每个单体有不同的 DNA 结合特异性，异二聚体识别的位点不同于同型二聚体。下面是对真核生物中几种 DNA 结合域的简单总结。

同型结构域蛋白质。同型结构域是一种典型的螺旋-转折-螺旋 DNA 结合域，其识别 DNA 的途径与细菌蛋白基本上是一致的(图 17-5)。不同蛋白质中同型结构域的结构非常相似，不仅识别螺旋相似，周边的把螺旋呈递给 DNA 的蛋白质结构也相似。相反，在前面的章节中可以看出，螺旋-转折-螺旋域的细节结构在很大范围内变化。在所有真核细胞中都发现了同型结构域蛋白质。如同在其他更高级的真核细胞中一样，果蝇中发现的这类蛋白质控制了许多基本的发育程序，我们将在第 19 章来讨论它们的功能。在酵母中也发现了同型结构域蛋白质，我们下面讨论的交配型控制基因就编码同型结构域蛋白质。图 17-5 显示了一种同型结构域蛋白质的结构。很多同型结构域蛋白质是以异二聚体的形式与 DNA 结合的。

含锌 DNA 结合域。各种不同的形式的 DNA 结合域包含锌原子。这包括典型的锌指蛋白（zinc finger protein）（如与核糖体 RNA 基因表达相关的通用转录因子 TFⅢA，第 12 章）和在酵母活化子 Gal4 中发现的相关的锌簇（zinc cluster）域。在这些例子中，锌原子同半胱氨酸残基和组氨酸残基相互作用，起到保持 DNA 结合域结构完整的作用（图 17-6）。DNA 同样被嵌入大沟中的 α 螺旋所识别。有些蛋白质拥有两个或多个末端相连的锌指结构域。每个域都以 α 螺旋嵌入大沟中，每增加一个锌指就延伸识别 DNA 序列的长度，以增强结合的亲和力。

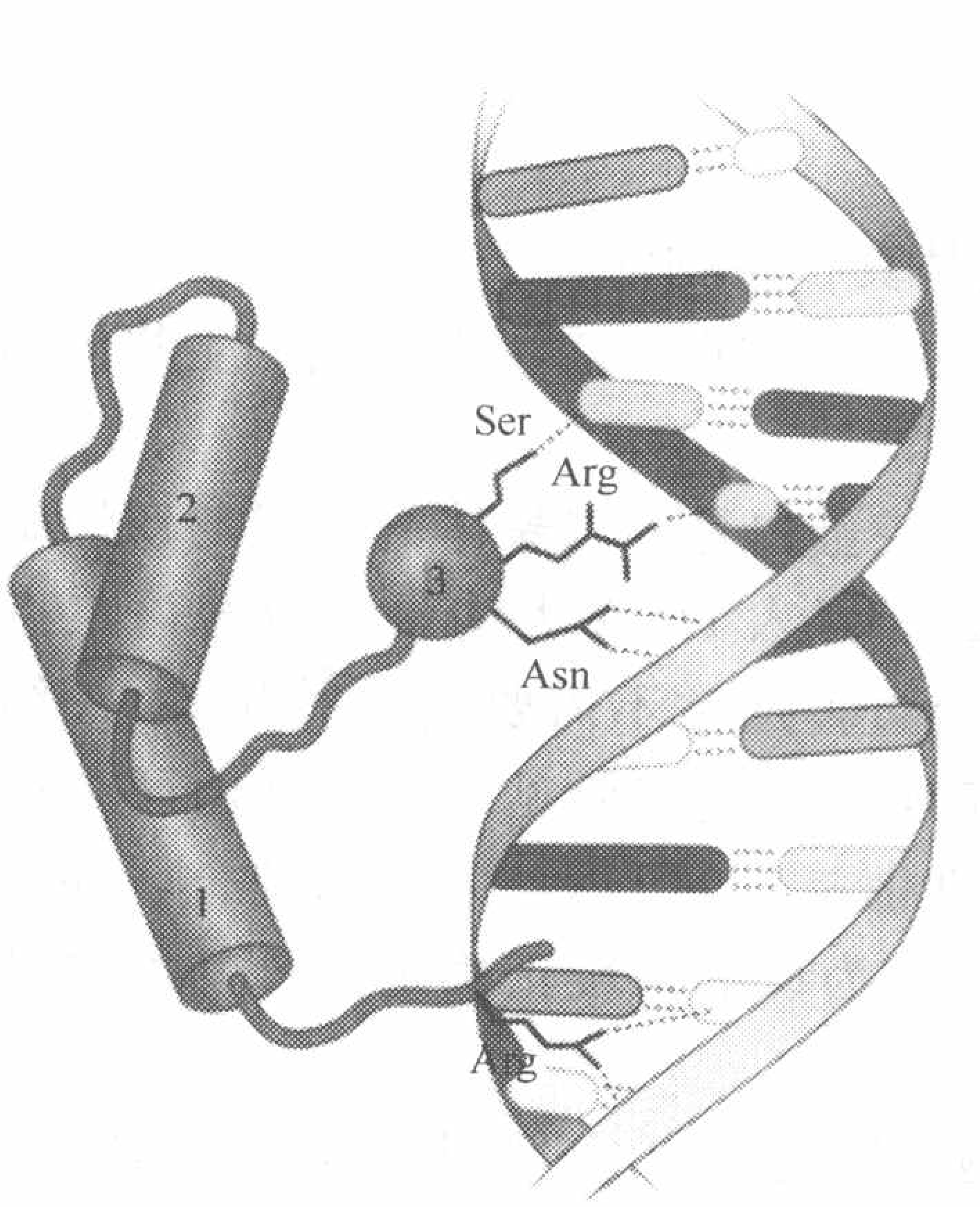


图 17-5　**同型结构域识别 DNA。**同型结构域由 3 段 α 螺旋组成，其中的两段（图中为螺旋 2 和螺旋 3）螺旋形成类似螺旋-转折-螺旋的模式（例如，将此图同图 16-2 相比较）。如此，螺旋 3 为识别螺旋，在图中它嵌入 DNA 大沟。沿着该螺旋外边缘的氨基酸残基同碱基对之间一一对应。在图中所示的例子中，酵母 α2 转录抑制子的螺旋 1 中伸出的一臂，同小沟中的碱基特异结合。（来源：Adapted with permission from Wolberger C. et al. 1991. Cell 67：517-528. © Elsevier）

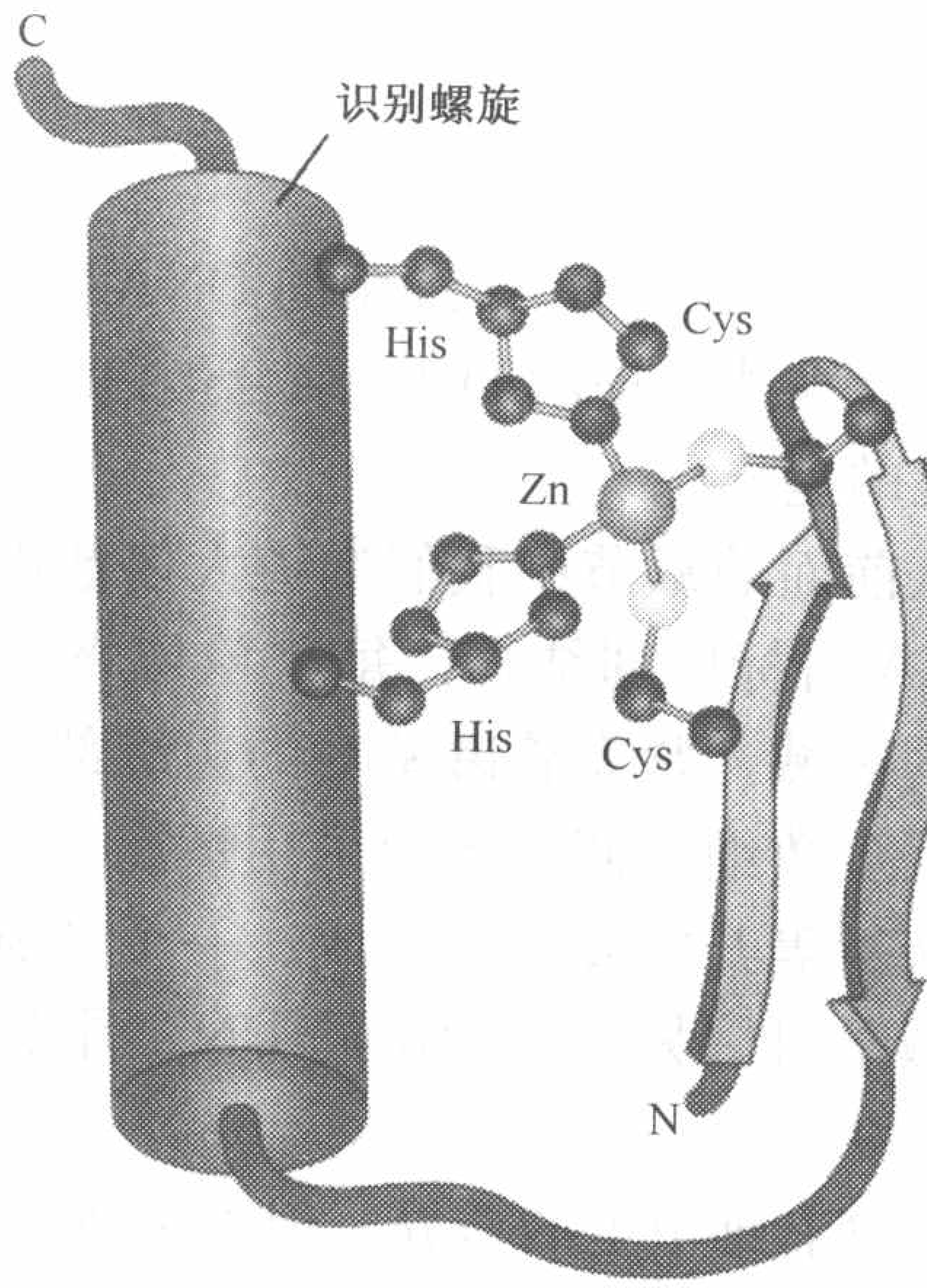


图 17-6　**锌指结构域。**结构左边 α 螺旋为识别螺旋，通过右边的 β 折叠伸向 DNA。如图所示，锌原子同 α 螺旋中两个组氨酸残基以及 β 折叠中两个半胱氨酸残基协同作用。这种排列有利于结构的稳定，是 DNA 结合所必需的。（来源：Adapted from Lee M. S. et al. 1989. Science 254：635-637.）

还有另一类使用锌的 DNA 结合域。在这些例子中，锌原子同 4 个半胱氨酸残基协同作用，稳定地形成一个类似于螺旋-转折-螺旋的 DNA 识别域。一个例子是在哺乳动物调节蛋白葡萄糖肾上腺皮质类脂醇受体中发现的，该蛋白质的作用是对某种激素信号做出反应，调节基因表达。

*亮氨酸拉链结构域。*该结构域是在单一的结构单元内同时包含二聚结构和与 DNA 结合的表面。如图 17-7 所示，两个长的 α 螺旋形成一个钳形结构，将 DNA 夹于其中，每个 α 螺旋嵌入大沟的一半转折部分。二聚化是由同样两个 α 螺旋内另一个区域介导的，在这个区域内，它们形成一个短的盘绕状结构，在这个结构中两条 α 螺旋通过适当间隔的亮氨酸残基（或其他疏水性氨基酸残基）之间的疏水相互作用结合到一起。我们已在第 5 章中讨论过这种蛋白质间相互作用的细节（图 5-15）。有亮氨酸拉链结构的蛋白质经常形成异二聚体，也形成同型二聚体。最后要提到的一类，即螺旋-环-螺旋蛋白（HLH 蛋白）也一样。

*螺旋-环-螺旋蛋白。*如同在亮氨酸拉链结构的例子中提到的一样，每个单体中延伸的 α 螺旋区域嵌入 DNA 的大沟中。如图 17-8 所示，二聚体的表面由两个螺旋区域构成，第一个是参与 DNA 识别的那个螺旋的一部分，另一个是一个较短的 α 螺旋。这两个螺旋被一个非刚性的环分离（因而得名），这个环容许两段螺旋靠近。由于与 DNA

结合的 α 螺旋区域内包含碱性氨基酸残基，亮氨酸拉链和 HLH 蛋白经常被称作碱性拉链（basic zipper）和碱性 HLH 蛋白(basic HLH protein)。

激活域结构不确定性

同 DNA 结合域相反，激活域一直没有一个明确的结构。在转录机器中，当与靶体之间相互作用时，它们呈螺旋结构，但可确信的是，这种结构是受 DNA 结合诱导的。我们将看到，缺乏明确的结构同激活域具有黏性表面，能同几种蛋白质的表面相互作用这个概念是相一致的。

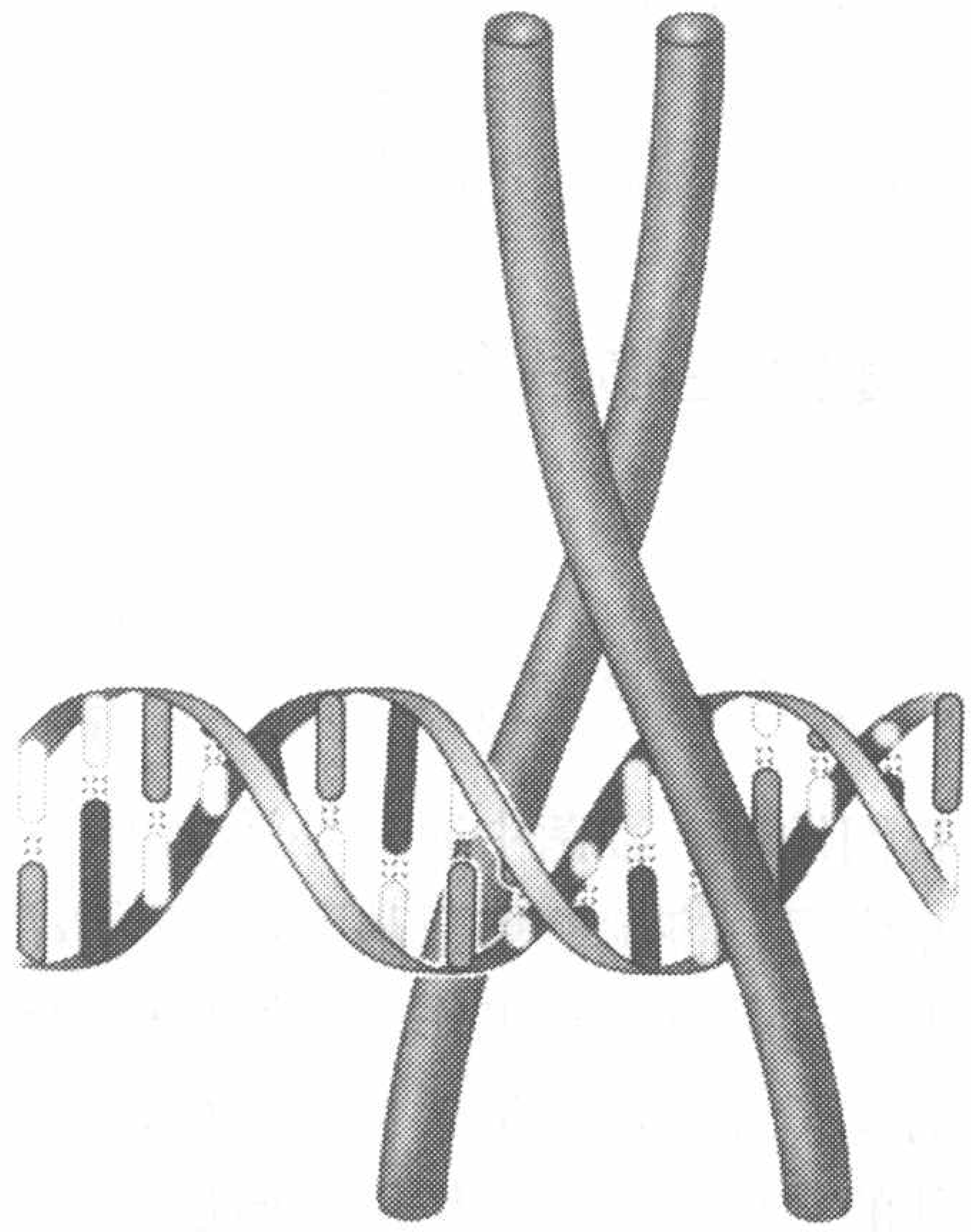


图 17-7 亮氨酸拉链结合 DNA。来自各自单体的两个巨大的 α 螺旋，在它们长链上不同的部分形成二聚体和 DNA 结合域。因此，如图所示，两个螺旋的顶部相互作用，形成盘绕状结构将两个单体结合在一起，往下，螺旋之间分离得足够开，以包含 DNA，并在 DNA 螺旋相对的位置嵌入大沟。另外，α 螺旋上氨基酸侧链与大沟中碱基对边缘之间的联系，提供了特异性。这个例子是在酵母转录活化子 GCN4 中发现的（图 5-15)。（来源：Adapted with permission from Ellenberger T. G. et al. 1992. Cell 71：1223. © Elsevier.）

不同于以结构来作为描述特征，激活域是按基本的氨基酸成分来归类的。例如，Gal4 的激活域被称作酸性激活域，反映出该区域酸性氨基酸占优势。突变实验凸现出了这些酸性残基的重要性，这些突变增强了活化子的效力：这样的突变总是增加整个激活域的酸性（负电荷）。除了这些，激活域还同样包含重要的疏水残基。很多其他活化子也具有同 Gal4 类似的酸性激活域，尽管在序列上看不到相似性，但是它们都保留了酸性的特征模式和疏水残基。

可以确信的是，激活域由许多重复的具有弱的活化能力的小单元组成，每个单元都是一段短的氨基酸序列。这样的小单元越多，每个单元的酸性越强，激活域的活化能力也就越强。这同激活域缺乏一个整体上的结构，只是简单地作为一种无差别的黏性表面在起作用这个概念是一致的。（为了理解这个推断，请想像用一个精确的、稳定的三维结构来代替折叠的激活域，类似于 DNA 结合域。在这样的环境下，不能指望 DNA 结合域的片段保留整个域的部分结合 DNA 的活性，完整的域结构是任何重要活性所必需的。但是，如果每个激活域都是简单的通用黏性表面，很容易想像它是由小的弱的单元组成。）

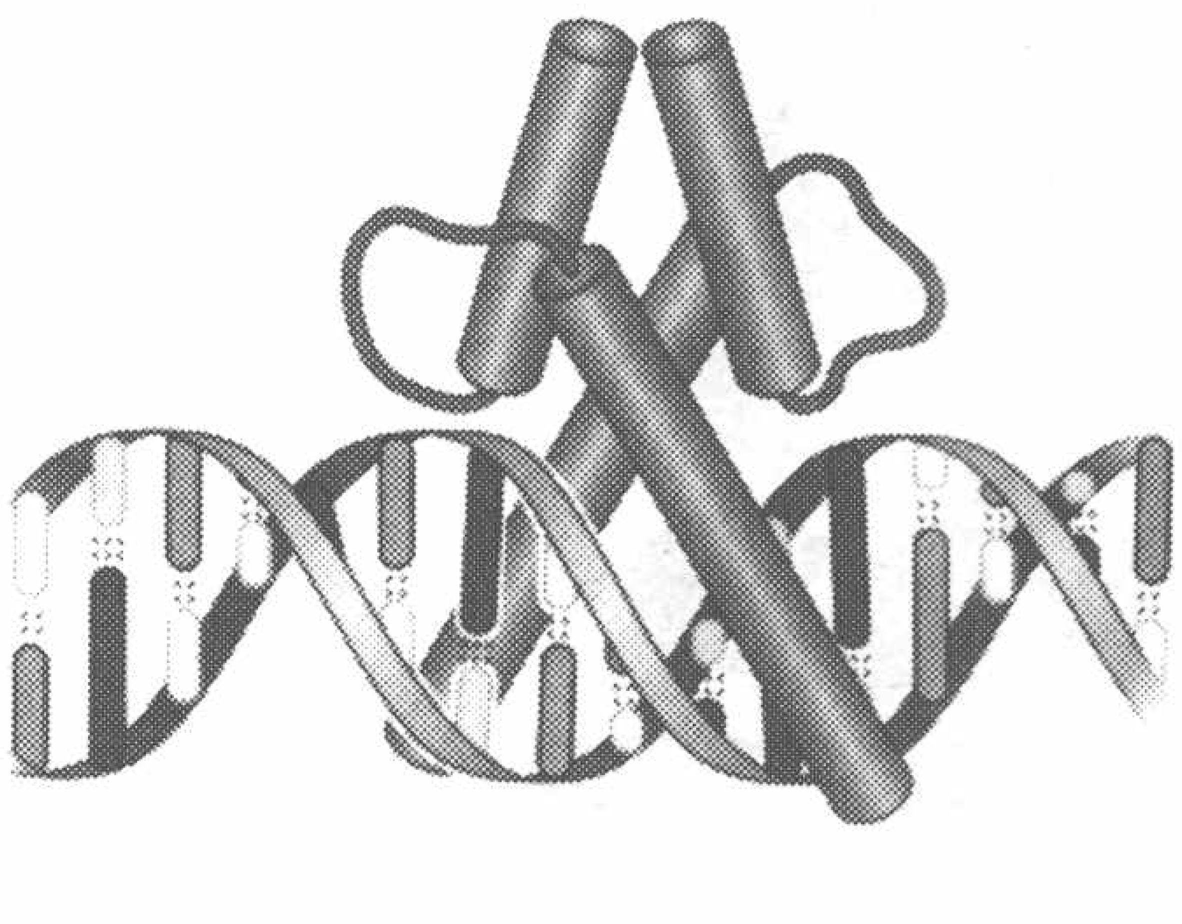


图 17-8 螺旋-环-螺旋模体。在这个例子中，我们再次看见一个长的 α 螺旋，参与 DNA 识别，并同一个短的 α 螺旋一起参与二聚体形成。（来源：Adapted with permission from Ma P. C. et al. 1994. Cell 77：451，Fig. 2A. © Elsevier.）

还有一些其他类型的激活域，包括谷氨酰胺

丰富的激活域，如在哺乳动物活化子 SP1 中发现的；脯氨酸丰富的激活域，曾在哺乳动物活化子 CTF1 中发现。这些功能域同样缺乏明确的结构。普遍来说，酸性激活域是常见的，并在所有检测过的真核生物中起作用，其他的激活域则很弱，也不是很常见。

通过真核生物活化子募集结合基因的蛋白复合物

活化子募集转录机器到基因上

我们发现在细菌中通常是活化子的一个表面与 DNA 结合，另一个表面与 RNA 聚合酶相互作用，将酶募集至基因，从而激活基因的转录（第 16 章，图 16-1)。真核生物的活化子也以这种方式来工作的，但是很少有活化子与 RNA 聚合酶之间直接相互作用的。相反，活化子或间接募集聚合酶或在聚合酶结合后募集其他因子。活化子与转录机器上聚合酶以外的其他部分相互作用，通过募集这些部分也就募集了聚合酶；另外，活化子能募集核小体修饰成分来改变基因附近染色质的性质，帮助起始转录。最后，活化子可以募集聚合酶进行转录起始或延伸所需要的因子。所有这些功能中，活化子只是将蛋白质募集到启动子上。在细菌中，RNA 聚合酶是唯一需要被募集的蛋白质；而在真核中则不同。在真核中，一个活化子能以 3 种方式起作用，我们首先来考虑转录机器的募集。

如第 12 章所提到的，除了 RNA 聚合酶，真核转录机器还包括很多其他蛋白质。这些蛋白质中的许多都形成复合体，如中介蛋白（mediator）和 TFⅡD 复合物（TFⅡD complex)（第 12 章的表 12-2 和图 12-16)。活化子与一个或多个复合体相互作用，将它们募集到基因上（图 17-9)。其他没有被活化子直接募集的成分，则同那些已被募集的成分结合，协同作用。

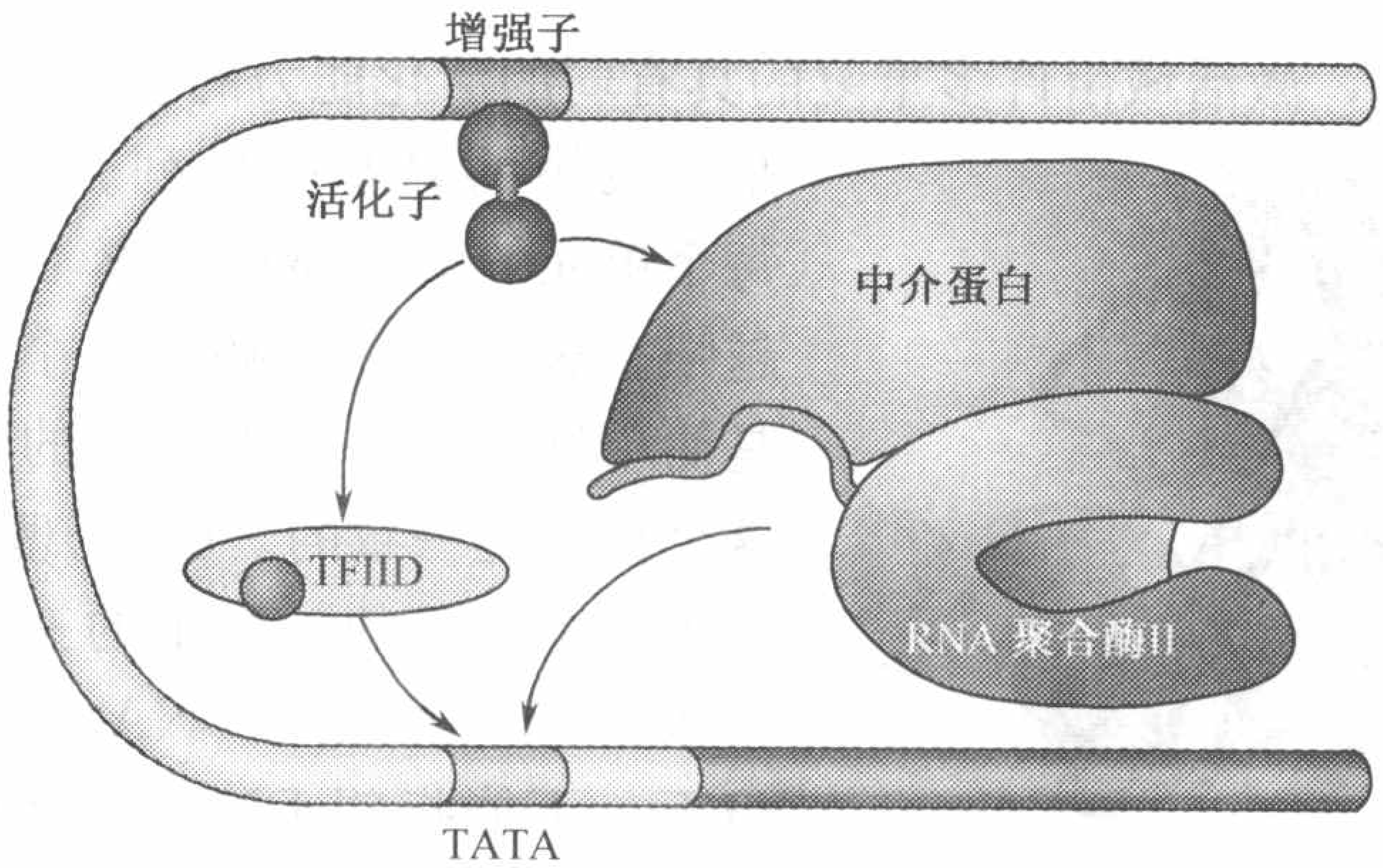


图 17-9 在真核细胞中通过募集转录机器来激活转录起始。图中，一个单一活化子募集两个可能的目标复合物：中介蛋白（通过其募集到了 RNA 聚合酶Ⅱ）和通用转录因子 TFⅡD。其他通用转录因子作为中介蛋白，RNA 聚合酶Ⅱ，或 IFⅡD 复合物的一个部分被募集，或单独被活化子直接募集，或在被募集的组分存在时与之随机结合，这些在图中没有显示出来。事实上，募集通常被多个结合在基因上游的活化子所介导。

转录机器中的许多蛋白能在体外同激活域结合。例如，一个典型的酸性激活域能够与中介蛋白的组分或 TFⅡD 的亚基相互作用。

募集可以通过染色质免疫沉淀反应（chromatin immunoprecipitation，ChIP）这种技术来观察（详见 21 章的描述）。该技术能揭示细胞内特定蛋白与 DNA 确定区域结合的时间。大多数基因的转录机器只在基因激活时才出现在启动子上，即转录机器不是预先装配好的，证实了活化子的作用是募集它。

在细菌中我们发现，通过募集来激活的基因（如 *lac* 基因）能通过一种活化子旁路实验来激活（框 16-1）。在这样一个实验中，当 RNA 聚合酶不经正常的活化子-聚合酶相互作用而募集至启动子时，也能观察到活化现象。在酵母中也有相似的实验，即 *GAL4* 基因（正常的应被 Gal4 激活）同样能被一种融合蛋白很好地激活，该蛋白质由细菌蛋白 LexA 的 DNA 结合域与中介蛋白复合物的一个组分直接融合而成（图17-10）。

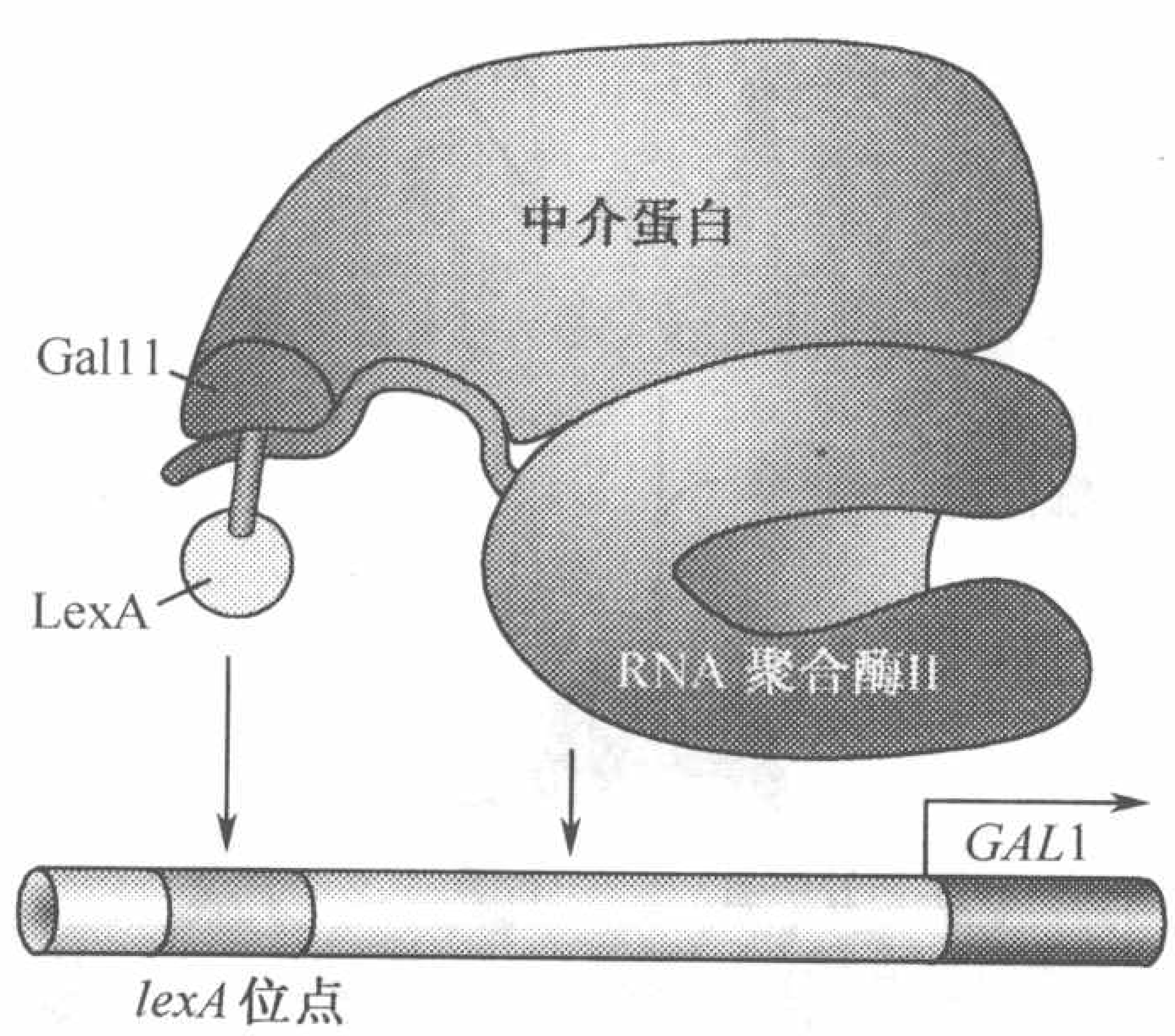


图 17-10 通过中介蛋白与 DNA 的直接联系来激活转录。这是一个第 16 章框 16-1 所描述的活化子旁路实验的例子。在缺乏通用活化子 Gal4 的条件下，*GAL1* 基因被一个融合蛋白所激活，该蛋白由 LexA 的 DNA 结合域与中介蛋白复合物的一个组分融合而成（Gal11/Med15，图 12-19）。活化依靠 LexA 的 DNA 结合位点嵌入基因的上游。其他转录起始所必需的组分（TFⅡD 等）假设同中介蛋白及 RNA 聚合酶Ⅱ结合在一起。

值得注意的是，这些实验没有排除这样一种可能性，即至少某些活化子不仅募集转录机器的部分组分，也诱导它们的变构，这样的改变也许能提高转录起始的效率。不过，将转录机器募集至基因是特异性的基础，即哪个基因被激活取决于哪个募集到转录机器。另外，活化子旁路实验的成功暗示在起始过程中变构事件的发生并不需要活化子做任何除了将蛋白质募集到基因上之外的事。

活化子也募集核小体修饰物来帮助转录机器结合到启动子上

除了直接募集转录机器，募集核小体修饰物也有助于激活染色质（chromatin）内那些难以接触到的基因。如同第 7 章讨论的，核小体修饰物有两种形式：一种是增加组蛋白末尾的化学基团，如增加乙酰基团的组蛋白乙酰转移酶（histone acetyl transferase，HAT）；另一种则重塑核小体，如依赖 ATP 活性的 SWI/SNF。这些修饰物是如何帮助基因活化的呢？有两个基本的模型解释了核小体的改变如何帮助转录机器与启动子结合（图 17-11）。

首先，重塑模型和某些修饰能够暴露 DNA 核小体内部的那些无法接近的结合位点。例如，通过增加核小体重塑因子的灵活性，提高了调控蛋白和转录机器的结合位点。相似地，组蛋白末端乙酰基团的增加，改变了组蛋白末端与邻近核小体的相互作用。这种改变常被说成松弛了染色质的结构，将这些位点释放（更完整描述参见第 7 章）。

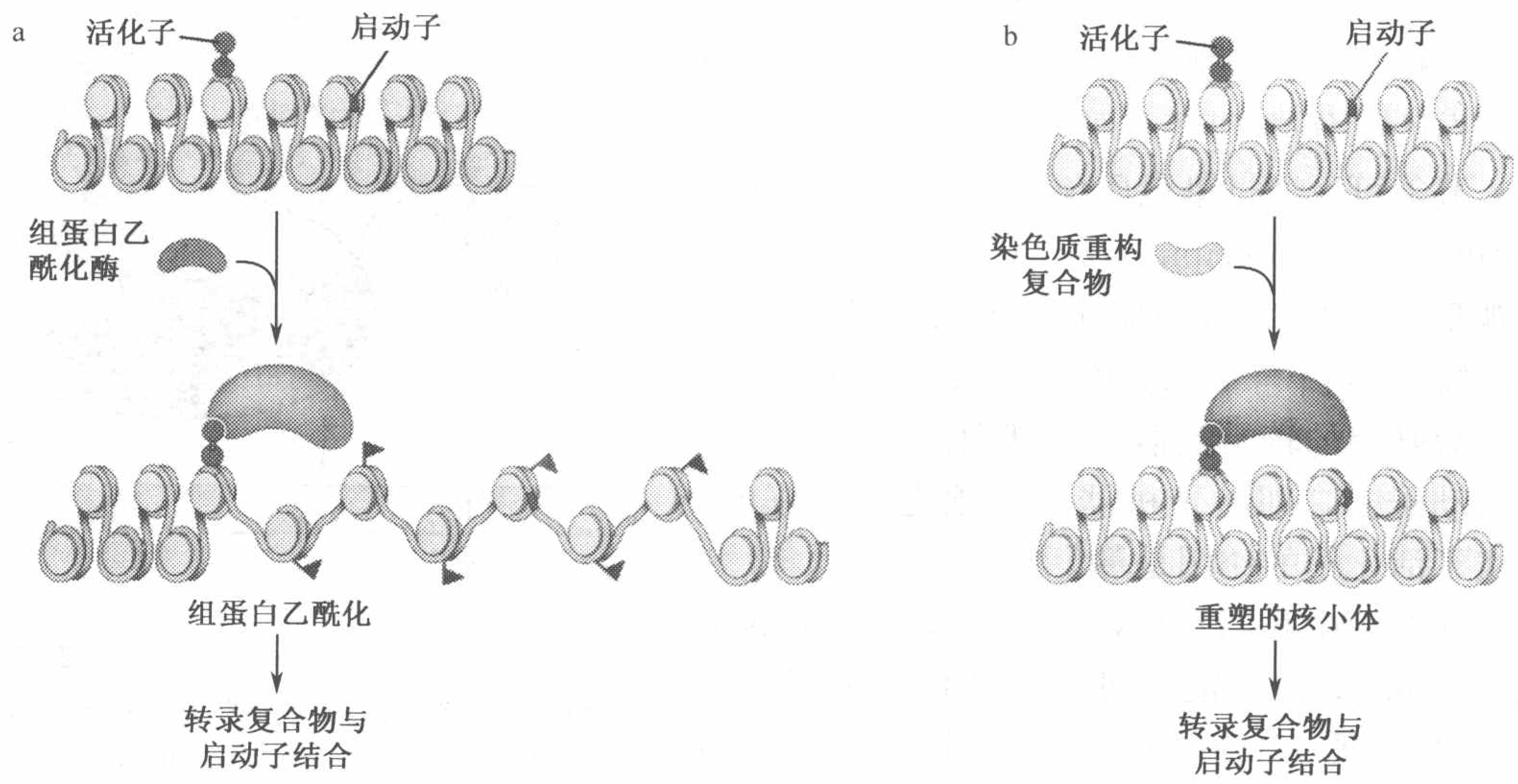


图 17-11 通过活化子直接改变染色质局部结构。在启动子上游边缘处所示的核小体内部的 DNA 上具有结合位点，由于不能接近染色质内部，活化子无法与这些位点结合。右图中，活化子募集某种核小体重塑分子来改变启动子附近的核小体结构，使之变得可接近并且能同转录机器相结合。左图中，活化子在募集组氨酸乙酰转移酶，该酶通过在组氨酸末端（如图中蓝色小旗所示）残基上添加乙酰基团来轻微改变核小体紧凑的结构，并为携带合适识别域的蛋白质创造结合位点（同源调节域；图 7-40 和图 7-41）。所有这些都介导了转录机器与启动子结合。

但是增加乙酰基团也以另一种方式协助转录机器（或其他蛋白）的结合：在核小体上为具有所谓同源调节域（bromodomain）的蛋白质提供特殊的结合位点（图 7-40）。TFⅡD 复合物的一个组分就具有同源调节域，与没有乙酰化的核小体相比，更易于与乙酰化的核小体结合。因此，若某个基因的启动子具有乙酰化的核小体，与那些没有乙酰化核小体的基因相比，将可能具有对转录机器更高的亲和力。

转录机器的哪一部分、哪种核小体修饰物是某个特定基因转录所必需的呢？哪个组分是被特定基因上特定活化子所直接募集的呢？在大多数情况中，这些问题的答案是未知的，但是转录机器的某些组分是某些基因更严格需求的，而其他基因就没那么严格，这同样适合于核小体修饰物。在许多事例中，这些差异并不是绝对的。因而，所有基因都绝对地需要 RNA 聚合酶，但是一个特定的基因也许依赖于转录机器另一个特别的组分，或者是某个核小体修饰物，或者什么都不需要。在某些例子中，也许只部分需要转录机器的一个组分（即缺少这个组分，活化效果会降低，但不会消除）。另外，一个特定基因激活的所需将随环境条件（如细胞周期的阶段）而变化，例如，在组蛋白乙酰基转移酶不存在时，Gal4 通常有效地激活 *GAL1* 基因。但是，有丝分裂时，当染色质更紧密时（第 7 章），活化效果将降低，除非基因募集了乙酰基转移酶。

在酵母中，最近的实验提供了在特异基因上特定的活化子与其目标之间相互作用的良好证据。看来，Gal4 与 3 个组分接触：中介蛋白、TFⅡD、和 SAGA（Spt-Ada-Gcn5-乙酰基转移酶）。其中最后一个组分含有乙酰化活性（表 7.8），似乎也能与转录

机器作用。事实上，与 TFⅡD 类似，SAGA 含有 TAFs（TBF 结合因子）（第 12 章），在某些启动子中用以取代 TFⅡD。像 Gal4 这样的酸性活化子作用于有着不同需求的基因（如，有些需要 TFⅡD 和中介蛋白，又有些需要 SAGA 和中介蛋白）的能力可以用其可与多个靶标相互作用的能力来解释。

我们注意到，一个 DNA 结合域与中介蛋白的一个亚基的融合蛋白足以激活转录（图 17-10）。尽管在这样的活化子旁路实验中激活达到了正常的水平，但是它缓慢；即需要花更长的时间使表达达到 Gal4 很快就能达到的水平。在活化子旁路实验中比较缓慢地达到激活的水平可能是因为在这种情况中需要转录机器的其他组分独立地抵达。

在某些启动子上活化子募集另一个因子来进行有效的起始和延伸

真核细胞中精细的转录机器含有很多用于起始的蛋白。它也含有一些帮助延伸的蛋白（第 12 章）。在某些基因中，启动子下游的序列导致聚合酶在起始后不久暂停或停滞。这些基因中，某些延伸因子的存在与否极大地影响了基因表达的水平。

一个例子就是果蝇的 *HSP70* 基因。这个基因由热激激活，是由两个活化子共同作用来控制的。GAGA 结合因子被认为将足够的转录机器募集到基因来进行转录起始。但是，如果缺乏第二个活化子 HSF，起始的聚合酶在启动子下游约 25bp 处停滞。应答热激，HSF 结合到启动子的特异位点，募集一个激酶 P-TEF（positive transcription elongation factor，正转录延伸因子）到停滞的起始机器上。该激酶磷酸化 RNA 聚合酶大亚基的羧基末端域（CTD 尾），释放停滞的聚合酶，使转录继续进行。

我们在第 12 章看到，在所有基因转录的早期，CTD 尾的磷酸化是一个重要的步骤，激酶 IFⅡH 可以进行这个磷酸化。是否大多数基因中都需要 P-TEF 还不清楚。像 Gal4 这样的强酸性活化子能够募集 P-TEF 以及转录机器的其他部分。可能只有在特定的一些基因中，不同的调控子在机器的募集中是像 *HSP70* 基因中看到的那样分开的，这使得基因转录增加了一次额外的控制。

导致 AIDS 的人类免疫缺陷型病毒（HIV）从由 P-TEF 控制的启动子转录其基因。可以再次看到，聚合酶在活化子 SP1 控制下从那个启动子起始转录，但不久后就停滞了。在这个例子中，P-TEF 被一个 RNA 结合蛋白而非 DNA 结合蛋白带到了停滞的聚合酶。那个蛋白叫 TAT。TAT 识别 HIV RNA 起始附近的一段特异的序列，该序列位于由停滞的聚合酶已经转录出来的部分。TAT 的另一个结构域与 P-TEF 相互作用，并将其募集到停滞的聚合酶。

远距作用：环与绝缘子

很多真核细胞活化子是远距作用的，尤其在高等的真核生物中。因此，在一个哺乳动物的细胞内，增强子也许能在它们所控制基因的上游（或下游）数十个甚至上百个 kb 的地方找到。我们发现在细菌中，结合在分离位点上的蛋白质仍然能通过 DNA 环的调节而相互作用。但是在这些例子中，我们只考虑相距数百碱基对的蛋白质结合位点。在这样的条件下，结合在 DNA 上的蛋白质相互之间充分靠近，它们之间在 DNA 上相互作用的机会比那些分离的蛋白质要高得多。一旦它们的结合位点之间的分离距离

超过几个 kb 时，这种优势将不复存在。

生物体存在着帮助远距离结合的蛋白质之间交流的机制。在细菌中，有一种途径可以做到这点，"建筑"蛋白 IHF（integration host factor）与 DNA 上的位点结合，并使之弯曲。在某些受 NtrC 控制的基因中，能在活化子结合位点和启动子之间发现 IHF 位点。通过弯曲 DNA，IHF 帮助与 DNA 结合的活化子到达启动子上的 RNA 聚合酶（图 16-4）。

多种模式被提出，以解释为什么在高等真核生物细胞中，结合在增强子和启动子之间的蛋白质能有助于活化作用。在果蝇中，*cut* 基因被 100kb 以外的一个增强子所激活。一种名为 Chip 的蛋白质（与同名的技术没有关系）有助于增强子和基因之间的交流。因此，Chip 蛋白的编码基因突变，影响活化的效力。Chip 蛋白是如何工作的，目前还是未知的，其中一个模型是该蛋白质同增强子和启动子间 DNA 上多个位点结合，通过多个 Chip 蛋白之间的相互作用，在中间的 DNA 上形成多个小环，这样的累积效果将使启动子和增强子接近。

在真核生物中还有其他几类模型，如我们所见，DNA 包在核小体中，受各种各样修饰的影响，这些核小体结构的部署、紧密程度和屈曲性发生改变。因此，被多个碱基对分开的位点也许在细胞中实际分开的距离并没有想像中那么远。而且在染色质某些位置也形成特殊结构，使增强子和启动子靠近。下一节，我们将描述几例远距离调控序列及其工作机制。但首先，我们来考虑远距离激活的另一个问题。

如果一个增强子能激活 50kb 外的特定基因，那么，是什么阻止它激活那些启动子在这个距离内的其他基因呢？一种名为绝缘子（insulator）的特殊元件控制了活化子的行动。当置于增强子和启动子之间时，绝缘子将抑制增强子对基因的激活作用。如图 17-12 所示，绝缘子并不抑制位于启动子下游的另一个增强子对同一基因的激活作用；同样它也不抑制最初的活化子对另一个基因的作用。因而，与绝缘子结合的蛋白质既不积极地抑制启动子的活性，也不抑制活化子的激活作用，它们只是阻碍着两者之间的联系。

在另一个实验中观察到绝缘子能抑制染色质修饰的展开。如我们所知，局部染色质的修饰状态能影响基因是否表达。如下所示，染色质上组蛋白修饰抑制的传播。是转录沉默（silencing）现象的核心。沉默是一种特殊的抑制形式，能沿着染色质展开，关闭多个基因的表达无需特殊的抑制子与基因结合。绝缘子元件能够阻断沉默的展开，所以绝缘子能使基因免于不加选择的活化作用和抑制作用。

这种情形对某些实验操作可以产生一些后果。由于与一种密度特别高的染色质（异染色质）结合，随机插入哺乳动物基因组的基因经常处于沉默状态，但是若在该基因的上下游放上绝缘子，该基因将不呈现沉默状态。

某些基因群的正确调控需要位点控制区域

人类珠蛋白基因在成人体内红细胞中和发育阶段那些形成红细胞的各种细胞中表达。人类有 5 种不同的珠蛋白基因（图 17-13a）。尽管这些基因聚集在一起，但它们并不全同时表达。相反，不同的基因在不同的发育阶段表达，首先表达的是 ε 基因，接着是 γ 基因，再后面则是 δ 基因和 β 基因。它们的表达是如何调节的呢？

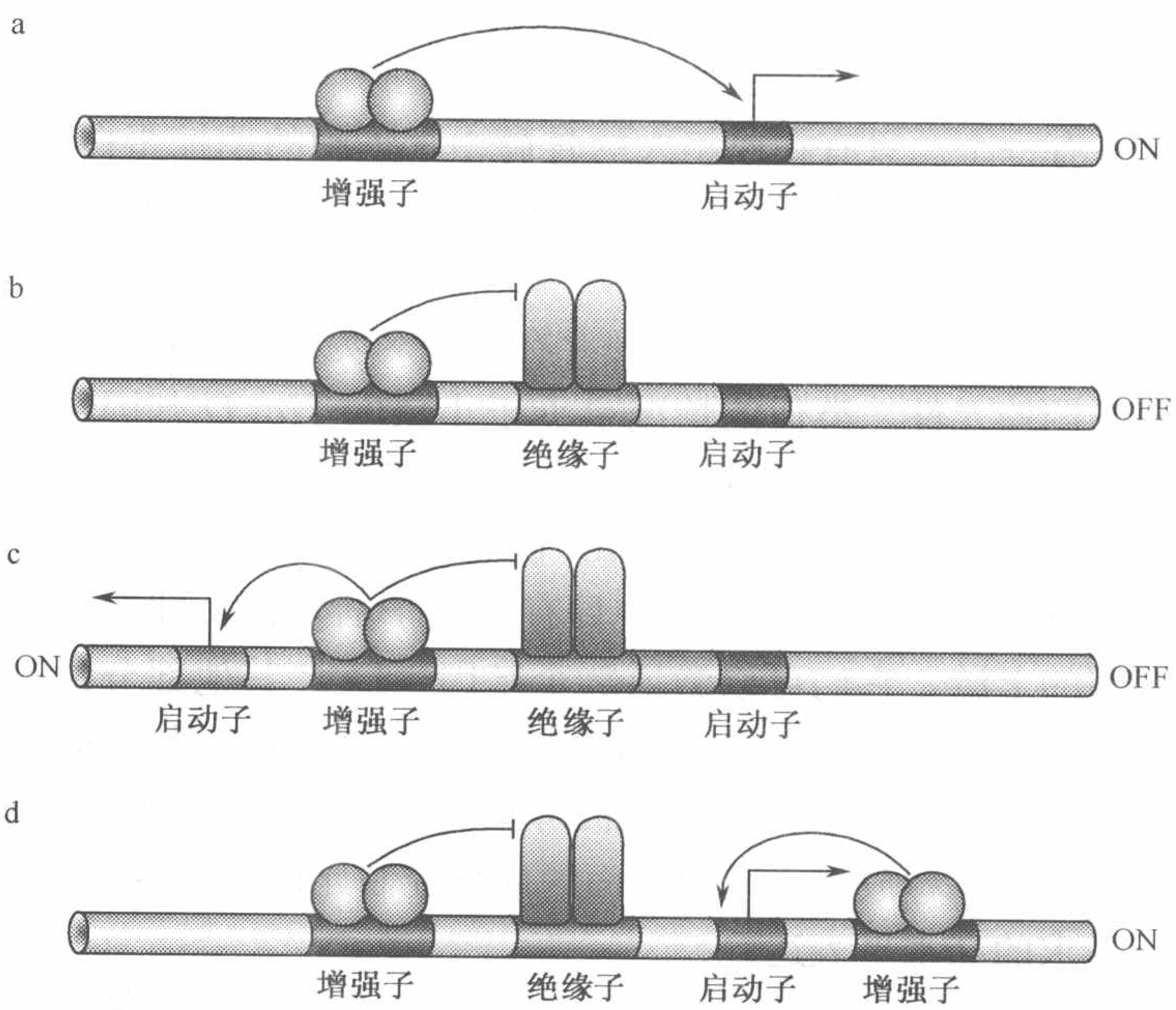


图 17-12　绝缘子阻止增强子的活性。a 与增强子结合的活化子对启动子的激活作用。b 一个绝缘子置于增强子和启动子之间。当适当的蛋白质与绝缘子结合后，尽管活化子与增强子结合，增强子对启动子的活化作用仍将被阻断。c 和 d 无论是在增强子上的活化子，还是启动子，都不会因为绝缘子的活动而失去活性。在 c 中，活化子能够激活附近的另一个启动子，在 d 中，起始的启动子能被下游的另一个增强子所激活。

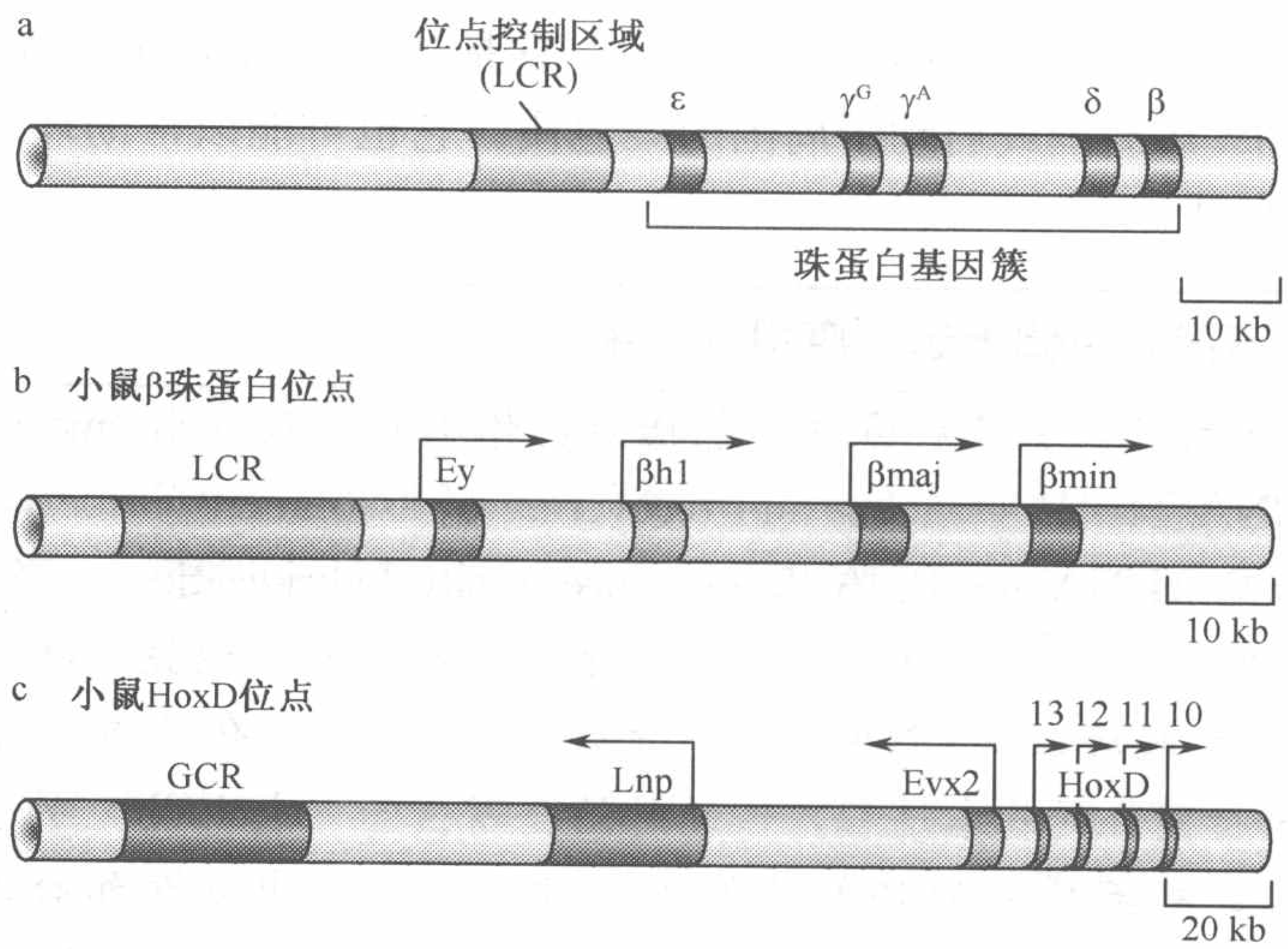


图 17-13　LCR 的调节。a. 人类珠蛋白基因及 LCR，后者确保了这些基因的有序表达。α 珠蛋白基因没有在图中显示，该基因的表达贯穿发育的始终，在不同的发育阶段，该基因的产物依次与图中所示的珠蛋白基因产物结合，形成不同形式血红蛋白。b. 小鼠的珠蛋白基因，同样受 LCR 的调节。c. 小鼠的 *HoxD* 基因簇，受一种名为 GCR 的元件的调节，该元件类似于 LCR，有强化基因簇有序表达的功能。

每个基因都有自己的一套调节位点，这些位点对于基因在发育阶段的特定时间和特定组织中的表达是必需的。这样，β 珠蛋白基因（在成年骨髓中表达）有两个增强子，启动子的上下游各有一个。只有在成年人的骨髓中，正确的调控蛋白才会全部起作用，并达到合适的浓度，与这些增强子结合。但是以正确的次序打开这些基因，还需要更多的东西。

在珠蛋白基因簇上游 30～50kb 处发现存在一组名为位点控制区域（LCR）（locus control region）的调控元件。类似的情况也见于小鼠的 *HoxD* 基因群。这些基因参与四肢的发育和斑纹化，在胚胎中以精确的方式表达（第 19 章）。*HoxD* 基因受一个类似于 LCR 的称为 GCR（global control region）的元件控制，工作机制类似于 LCR。

LCR 由多个序列元件组成。其中某些具有增强子的特征：即在实验中将这些序列置于一个报道基因的上游，它们可以激活那个基因。LCR 其他部分的行为更像绝缘子元件，还有一些部分看上去则具有启动子的特征。元件多样性导致多个 LCR 作用机制模型的产生。一个模型提议整个转录机器被募集到 LCR，从那里转录整个位点，随着机器的前行打开染色质，释放每个基因前面的局部控制元件。这些单个的启动子随后将产生每个基因所需求的高水平表达。

最近的实验使用新的技术，可以使活化过程中 LCR 和启动子在细胞中的定位可视。已经对许多受 LCR 或类似 LCR 元件控制的基因进行了该项实验。所有的结果都显示，当一个启动子被激活时，结合于上游调控序列的调控蛋白被发现位于该启动子的附近。这个结果与如下想法一致，即结合在 LCR 的蛋白质与结合在启动子上的其他蛋白质相互作用，居间的 DNA 形成环状突起以使这种相互作用得以进行。关于远距离机制如何工作的更详细内容，参见框 17-2。

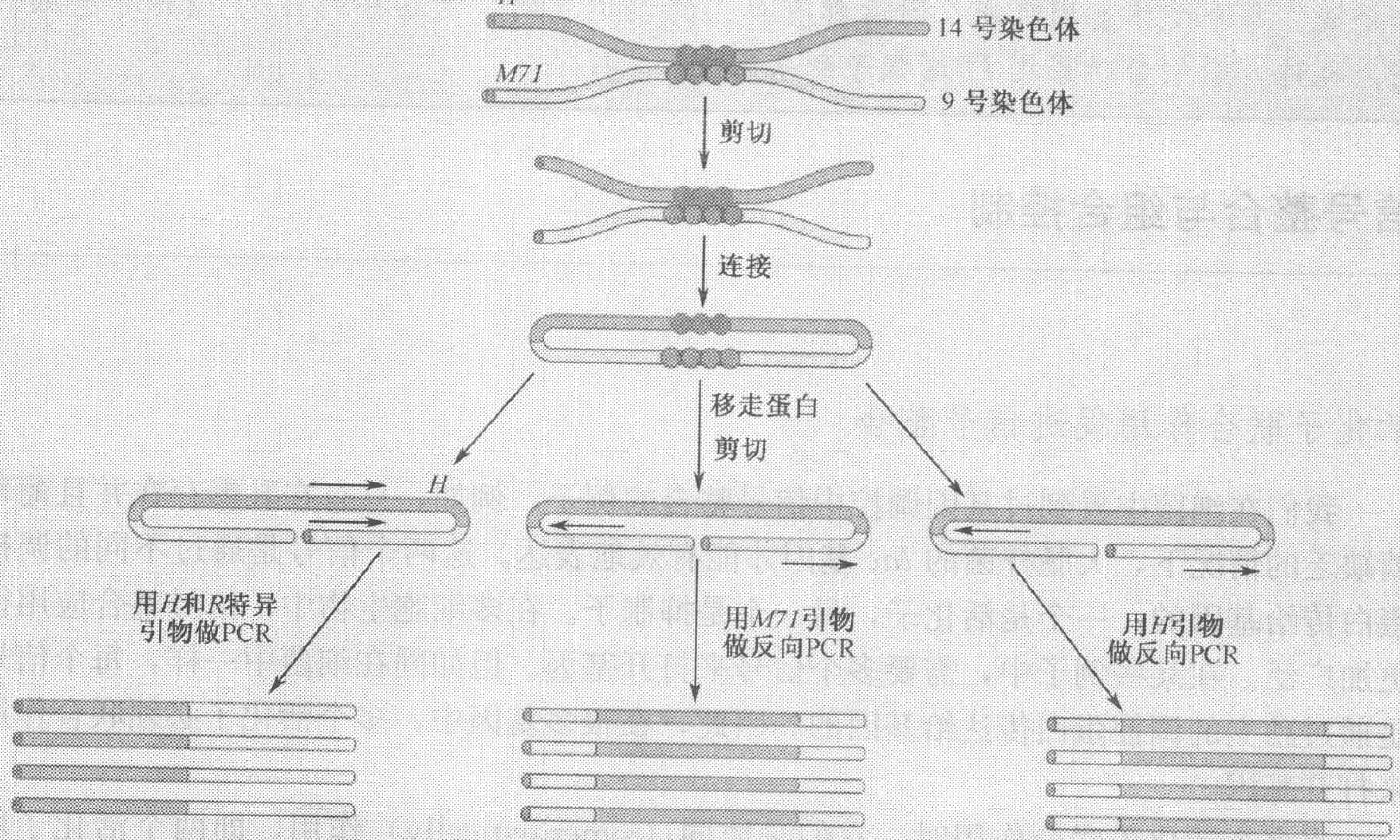
LCR 的激活与丰富的核小体修饰相关联。它是如何与激活挂起钩来的还不清楚。它可能帮助打开 LCR 周围或启动子周围的染色质。它也可能以帮助环形成的方式来改变两者之间的染色质。

框 17-2　同一和不同染色体上远距离相互作用

LCR 和启动子共定位有各种方法，包括染色体构象捕获（chromosome conformation capture，3C），RNA-TRAP（tagging and recovering of associated proteins），以及 DNA 和 RNA 荧光原位杂交（DNA and RNA fluorescence in situ hybridization，FISH）。

第一个方法是第 21 章将讨论的 CHiP 技术的一个变种，该技术将蛋白交联到其结合的 DNA 位点。交联后，将 DNA 断裂，随后可以确定任何仍然与蛋向结合的短片断。3C 方法中，也进行同样的基本过程，但此时结合增强子（或 LCR）的蛋白质不仅交联 DNA，也交联与之相互作用的任何其他蛋白质。如果其中包含结合其他 DNA 位点（如，启动子）的蛋白质，那么它们的 DNA 位点也可以被确定，如框 17-2 图 1 所示。

还有其他远距离激活的例子，调控序列不仅离它们所控制的启动子很远，而且在另一条染色体上！这些情况也被上述提到的技术检测到。例如，小鼠 γ-干扰素基因位点（10 号染色体上）的调控区域与 11 号染色体上编码 T 辅助细胞 2 细胞因子基因的位点相互作用，已知在这些细胞中后者的表达与干扰素基因的表达相关联。



框 17-2 图 1　染色体构象捕获揭示了 *H* 增强子与嗅觉受体基因的关联。 图解显示 3 个不同的 3C 策略，揭示了 *H* 增强子元件与 *OR* 启动子的关联。左图，用了一个 *H* 特异的聚合酶链反应（polymerase chain reaction，PCR；见第 21 章）的引物和一个 *OR* 启动子特异的引物。中图，用了一对 *M71*（编码小鼠嗅觉受体）特异的反向引物来分离与 *M71* 基因关联的序列。右图，用了一对 *H* 特异的反向 PCR 引物来检测与 *H* 元件共定位的 DNA 序列。（来源：Adapted from，with permission，from Lomvardas et al. 2006. *Cell* 126：403-413，Fig. 1A. ©Elseveir.）

更神奇的也许是小鼠感官神经细胞中嗅觉受体（odor receptor，*OR*）基因的激活。在每个神经细胞中有 1300 个受体基因，但每个神经细胞只表达其中一个。另外，不管哪个基因被选中，其表达都由同一个增强子（*H*）来激活。

使用 3C 技术，最近发现，在单个神经细胞中，*H* 增强子总是被发现位于该细胞中表达的那个特定的基因的启动子附近，无论该基因是否与增强子在同一条染色体上。这个观察结果产生了一个模型，即结合在增强子上的活化子在一个特定的细胞中随机地选择一个嗅觉受体基因（假定是它第一个遇到的随便哪个基因）。增强子只有一个拷贝的事实确保了用这种方法只有一个基因可以被激活。支持这个模型的证据包括这样一个实验，人工加入额外拷贝的增强子后，产生的神经细胞每个可以表达多于一个的嗅觉受体。

染色体间反式激活的例子先前在果蝇中有报道。在大多数这些例子（称为基因转应作用，transvection）中，一个基因的一个同位基因被与其他同位基因（位于姐妹染色体上）关联的增强子激活。于是人们认为两个元件是通过染色体的同源配对而被带到一

起的。在像嗅觉受体这样的例子中，还不清楚增强子和目标基因是如何相互找到的。有 1300 个可能的基因供选择的事实降低了找到某一个的低效性，但是作用机制仍然还是一个谜。每个受体基因可能在附近都有自己的调控序列，这些序列与 *H* 增强子一起作用。也许，它们帮助募集 *H* 增强子到基因上。

信号整合与组合控制

活化子联合作用促进信号整合

我们在细菌中看到过基因调控中信号整合的例子。例如，只有在乳糖存在并且葡萄糖缺乏的情况下，大肠杆菌的 *lac* 基因才能有效地表达。这两个信号是通过不同的调控蛋白传给基因的，一个是活化子，另一个是抑制子。在多细胞生物中，信号整合应用得更加广泛。在某些例子中，需要多个信号来打开基因。但如同在细菌中一样，每个信号是通过独立的调控蛋白传达给基因的。因此，在很多基因中，多个活化子必须联合作用来打开基因。

当多个活化子联合作用时，它们是协同（synergistically）作用，即两个活化子联合作用的效果要好于（通常是非常明显的）每个活化子单独作用的效果之和。协同有 3 个来源：多个活化子各募集转录机器的同一组分；多个活化子各募集转录机器的不同组分；多个活化子相互之间帮助与所调控基因上游的位点相结合。在给出具体的事例之前，我们简述这 3 种策略如下。更多的转录协同例子见框 19-5。

通过与不同的部位接触，两个活化子能募集同一复合体，如中介蛋白。组合的结合能对募集有指数效果（表 3-1）。在那些活化子募集不同复合物（若没有帮助，两者都不能有效地与 DNA 结合）的例子中，协同效应更易于被描述。

当一个活化子与 DNA 的结合依赖于另一个活化子与 DNA 的结合时，活化子能相互帮助与 DNA 结合，协同同样也来自于此。协作（cooperativity）可以是我们在细菌中遇到的那种，即当与 DNA 位点结合后，两个活化子由此相互接触（如图 16-25 所示的 λ 抑制子的情形）。但是这也能以其他方式起作用：一个活化子募集某些物质来促进第二个活化子与 DNA 结合。图 17-14 阐明了活化子之间通过相互帮助与 DNA 结合的各种方式。这包括“经典的”协作结合，一个活化子募集修饰物来帮助另一个活化子结合 DNA，一个活化子与核小体 DNA 结合使另一个活化子的结合位点暴露出来。

协同对于活化子的信号整合至关重要。有一个基因它的产物只有当两个信号都接收到时才需要。每个信号是通过单独的活化子传达给基因，只有当两个活化子都存在时，基因才能有效地表达，而任何一个活化子的单独作用将不能影响基因的表达。

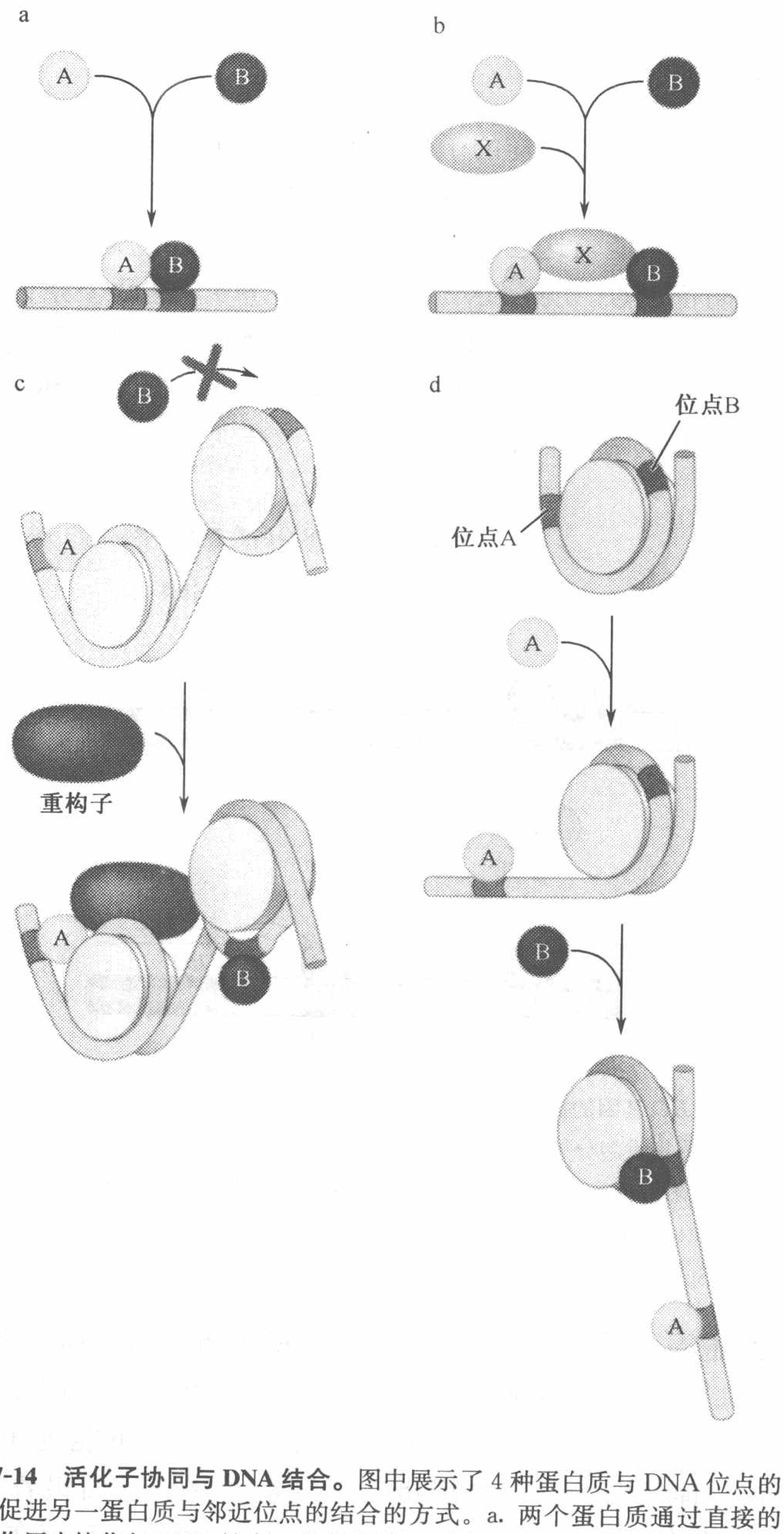


图 17-14 活化子协同与 DNA 结合。图中展示了 4 种蛋白质与 DNA 位点的结合促进另一蛋白质与邻近位点的结合的方式。a. 两个蛋白质通过直接的相互作用来协作与 DNA 结合。我们在第 16 章的 λ 抑制子中能看到这种方式，在很多真核细胞调控蛋白中也有这种方式。b. 两个蛋白质同共有的第三个蛋白质作用来实现类似的作用。c、d 蛋白质与核小体内 DNA 位点的结合促进另一蛋白质结合的间接方式。c. 第一个蛋白质募集核小体重塑分子来使第二个蛋白质的结合位点暴露。d. 第一个蛋白质与它的位点结合，这个 DNA 位点恰巧在核小体外。结合后稍微展开了的核小体 DNA，暴露第二个蛋白质的结合位点。所有这些机制都能解释为什么一个调控蛋白能促进其他调控蛋白的结合，或者更进一步一个活化子如何促进转录机器与启动子结合。

信号整合：*HO* 基因受两个调控子的控制，一个募集核小体修饰物，另一个募集中介蛋白

不同于通过分裂产生两个同样的子代细胞，酿酒酵母的分裂方式为出芽分裂，即所谓的母细胞通过出芽产生一个子代细胞。在此我们将集中讨论 *HO* 基因的表达。（我们不需要关注这个基因的功能，这已在第 11 章中阐述。）*HO* 基因仅在母细胞内表达，并且只出现在细胞周期特定点（见演示 17-1 和第 19 章）。这两个条件是通过两个活化子（SWI5 和 SBF）传达给基因的。SWI5 在远离基因处有多个结合位点，最近的一个离启动子也有 1kb 以上(图 17-15)。SBF 也有多个结合位点，但它们都离启动子很近。为什么基因的表达要依赖于这两个活化子呢？

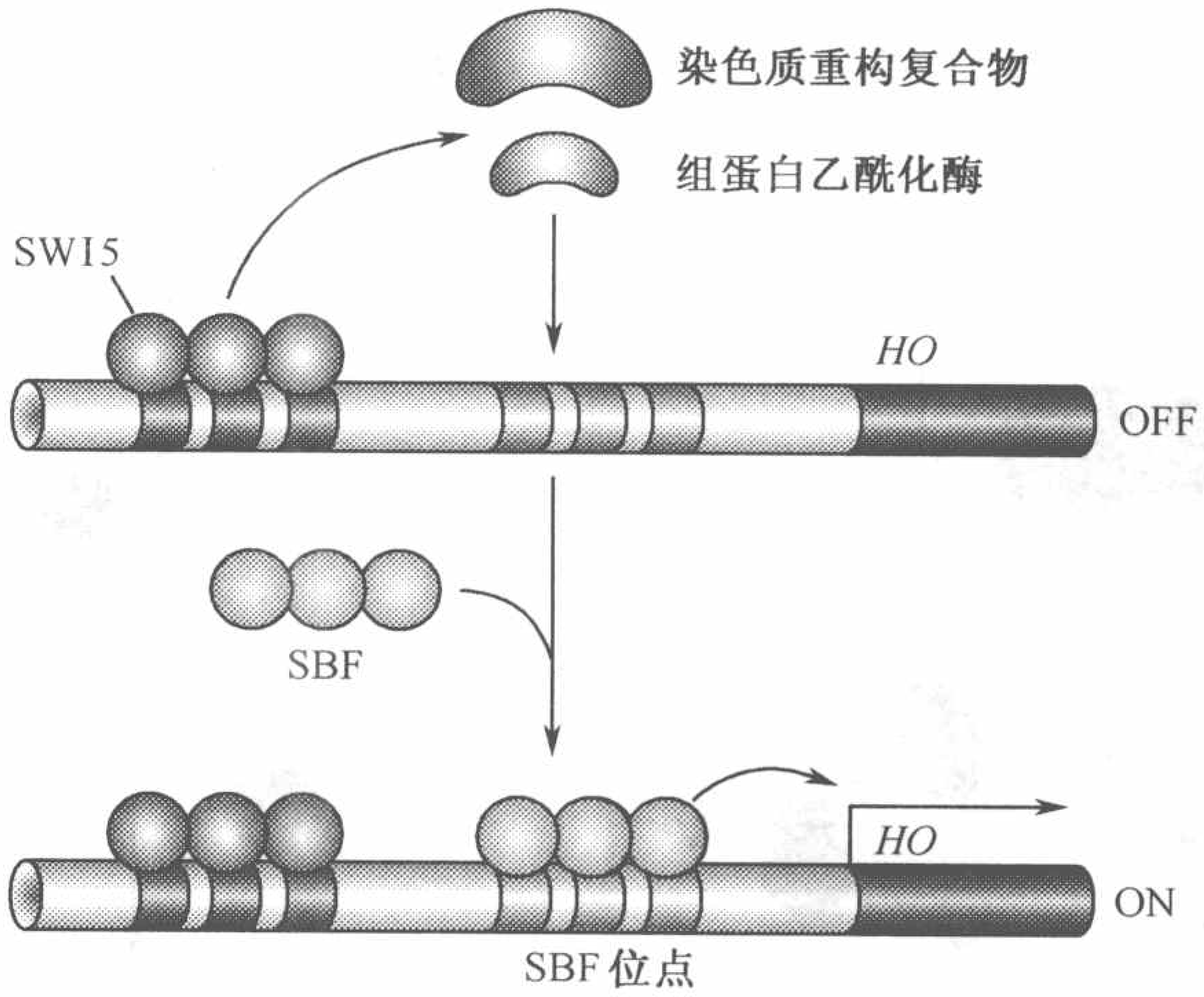


图 17-15　对 *HO* 基因的控制。在染色质中，SWI5 能独立地与它的位点结合，而 SBF 不能。SWI5 募集重塑分子和组蛋白乙酰基转移酶改变核小体，暴露 SBF 的位点，使得 SBF 在启动子附近结合，从而激活基因。在本图中，为了简便起见，没有画出核小体。（来源：Adapted from Ptashne M. and Gann A. 2002. Genes & Signals，p. 95，Fig. 2-18. © Cold Spring Harbor Laboratory Press.）

SBF（仅在细胞周期适当阶段具有活性）在染色质中的排列使之不能独立地与位点结合，SWI5（仅在母细胞内有活性）能独立地与位点结合，但是在那样一个距离上（大于 1kb）不能激活 *HO* 基因（记住在酵母中，活化子不能远距离起作用）。然而，SWI5 能够募集核小体修饰物（一种组蛋白乙酰基转移酶后面跟着重塑酶 SWI/SNF）。其在核小体上的作用将暴露 SBF 位点。这样，若两个活化子都存在，并且都具有活性，SWI5 的行为使 SBF 能与位点结合。接着，通过直接与中介蛋白结合，SBF 能募集转录机器并激活基因的表达。

信号整合：活化子协同作用，共同与人类 β 干扰素基因结合

病毒感染能够激活细胞内人类 β 干扰素基因。感染触发了 3 个活化因子：NFκB、IRF 和 Jun/ATF。在人类 β 干扰素基因启动子上游约 1kb 处有一个增强子，

3个活化子将协同地一个挨着一个与该增强子结合。这些活化子以协同作用结合在增强子上形成的结构被称作增强体（enhanceosome）（图17-16）。活化子随后募集一个所谓的共活化子，一个称为CBP（CREB-binding protein）的蛋白或其近亲p300。这个蛋白具有组氨酸修饰活性，能够募集核小体重塑活性（如SWI/SNF）及转录机器本身。

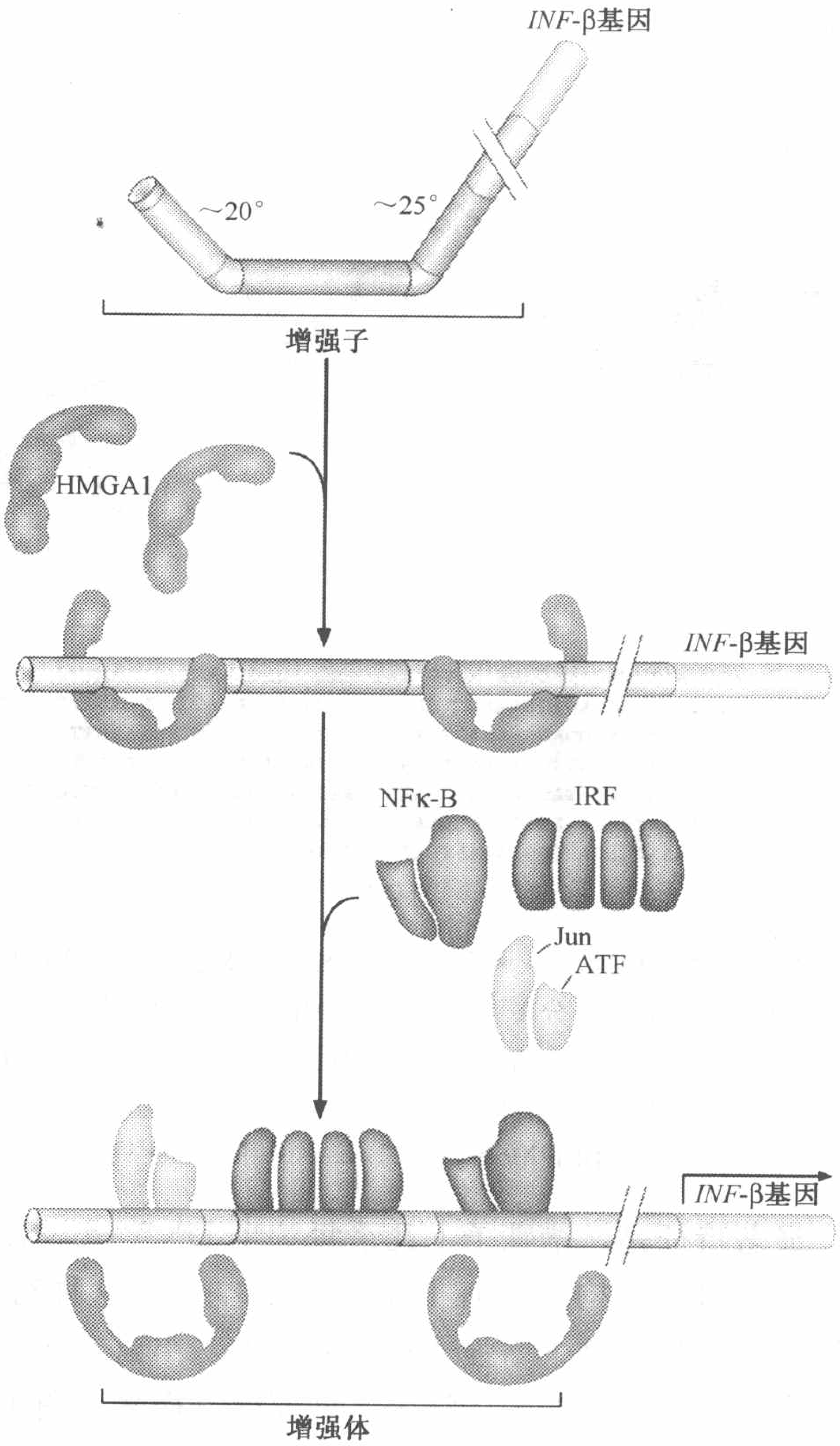
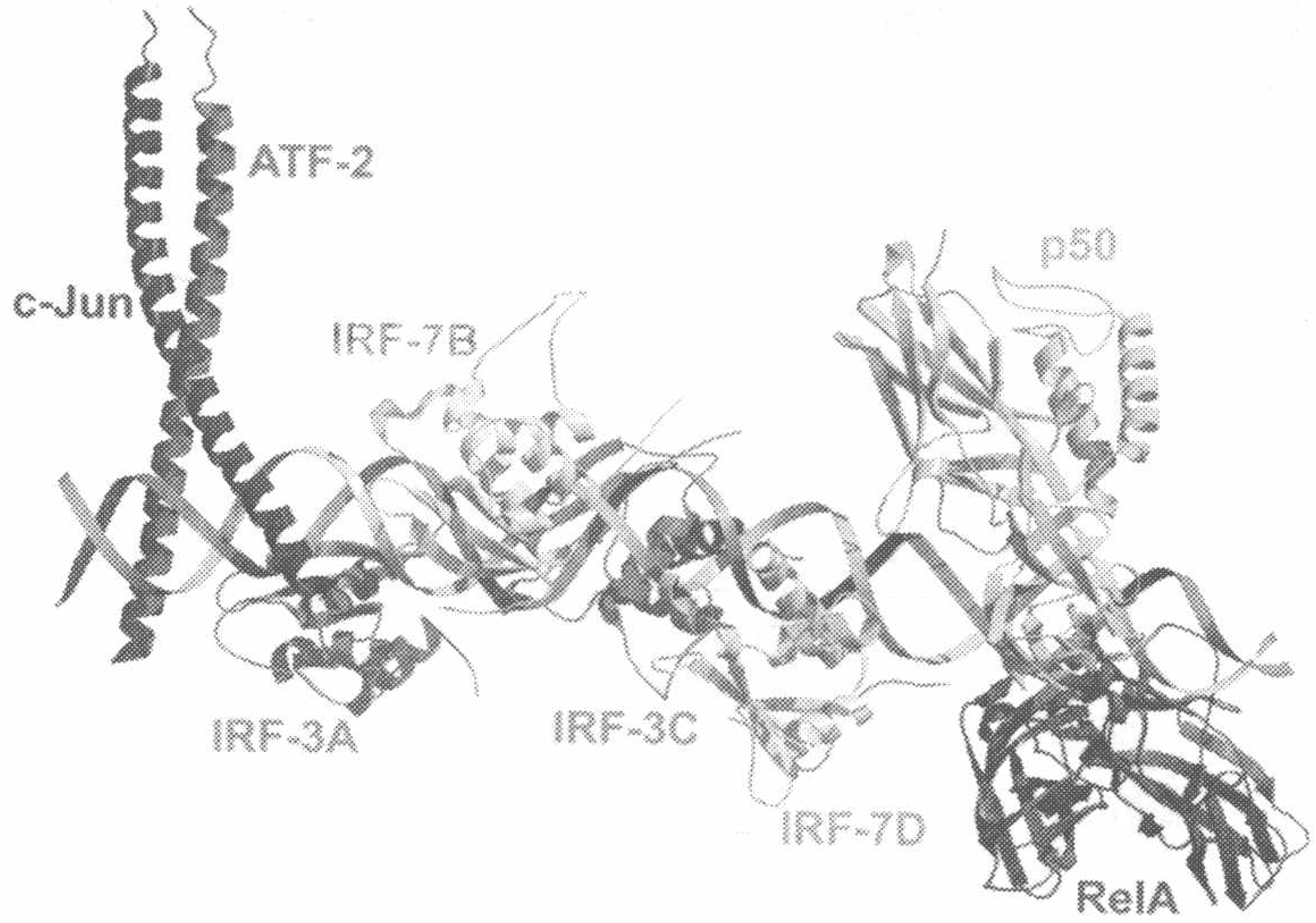


图17-16 人类β干扰素增强体。 3个活化子的协同结合，与结构蛋白HMGA1一起激活β干扰素基因。

除了上述的活化子，另一个蛋白HMGA1结合增强子。这个蛋白结合在DNA反面的小沟里，并帮助增强体的组装，尽管它可能不是最终结构的一部分。事实上，一旦所有的活化子都结合上去，它看上去无法保留在上面；很简单，DNA上太拥挤了。

图 17-16 显示人们认为的增强体组装方式，图 17-17a 显示结合在增强子 DNA 上的所有活化子的 DNA 结合域的晶体结构。如图 17-16 所示，增强子 DNA 是弯曲的，但一旦活化子结合了，它就变直了；HMGA1 撑直了 DNA，因而帮助最终结构的形成。

a



b

```
              ATF  Jun  IRF   IRF   IRF   IRF    NF-κB
Human     1:AAATGTAAATGACATAGGAAAACTGAAAGGGAGAAGTGAAAGTGGGAAATTCCTCTGAAT:60
Mouse     1:......AAATGACAGAGGAAAACTGAAAGGGAGAACTGAAAGTGGGAAATTCCTCTGA..:52
Rat       1:......AAATGACGGAGGAAAAGTGAAAGGGAGAACTGAAAGTGGGAAATTCCTCTGA..:52
Swine     1:......AAATGACATAGGAAAACTGAAAGGGAGAACTGAAAGTGGGAAATTCCTCTGAA.:53
Horse     1:.AATGTAAATGACATAGGAAAACAGAAAGGGAGAACTGAAAGTGGGAAATTCCTCTGAA.:58
Bovine2   1:.....TAAATGACAAAGGAAAACTGAAAGGGAGAACTGAAAGTGGGAAATCTCTCC....:45
Bovine    1:.....TAAATGACATGGGAAAAATGAAAGCGAGAACTGAAAGTGGGAAATTCCTCT....:51
```

1亿年

图 17-17 增强体结构及序列。 a. 增强体的晶体结构，揭示结合在增强子 DNA 上的活化子的 DNA 结合域。b. 分离了 1 亿年的物种间 β 干扰素增强子 DNA 序列的保守性。还显示了增强子中每个活化子识别的序列。（来源：a，Panne D. et al. 2007. *Cell* 129：1111. PDB Codes：2061，206G.）图像是多结构的合并，由 Leemor Joshua-Tor 组装成模型。图像用 MolScript，Bobscript 和 Raster 3D 制作。

这个结构的一个显著特征是，增强子上几乎每一个 DNA 碱基对都参与活化子的结合，这就是为什么认为在最终结构里没有 HMGA1 的空间。这种对增强子中序列信息的彻底使用也可能解释了为什么这个增强子在跨越像人、小鼠、马这样不同的物种中如此高度地保守；事实上，它甚至比基因的编码序列要保守得多（图 17-17b）。

如我们已经注意到的，活化子结合（及增强体形成）以一种高度协同的方式进行，确保所有 3 个活化子都必须在场。下面是调控子可能协同结合的 3 种方法：(1) 它们之间的蛋白-蛋白相互作用；(2) 一个蛋白结合引发的 DNA 改变帮助另一个蛋白的结合；(3) 所有的活化子同时与共活化子 CBP 相互作用。所有 3 种方法可能在这个例子中都使用到，尽管要知道活化子之间的蛋白-蛋白相互作用的程度是非常困难的。在这个结构中看不出多少直接相互作用的证据，不过，多数情况下在结构中只显示活化子的 DNA 结合域。

组合控制是真核细胞的复杂性与多样性的核心所在

在细菌中，我们遇到过简单的组合控制（combinatorial control）。如 CAP 能通过与不同调控蛋白的协作来调节多个基因。在 *lac* 基因中，CAP 同 Lac 抑制子共同作用，而在 *gal* 基因中，它又同 Gal 抑制子协作（第 16 章）。

在真核细胞中存在着广泛的联合控制。我们首先考虑一个常见的例子（图17-18），基因 A 受 4 个信号（分别为 1、2、3、4）的控制，每个信号通过一个独立的活化子（即活化子 1、2、3、4）来起作用。基因 B 受 3 个信号（分别为 3、5、6）的控制，这些信号通过对应的 3 个活化子来起作用。注意，信号 3 是在两个基因中都有的信号，活化子 3 同样作用于两个基因。在复杂的多细胞生物（如果蝇和人类）中，联合控制涉及比上例中更多的调控蛋白和基因，当然，抑制子示例活化子都可能参与。调控蛋白何以能移如此杂乱地混在一起？

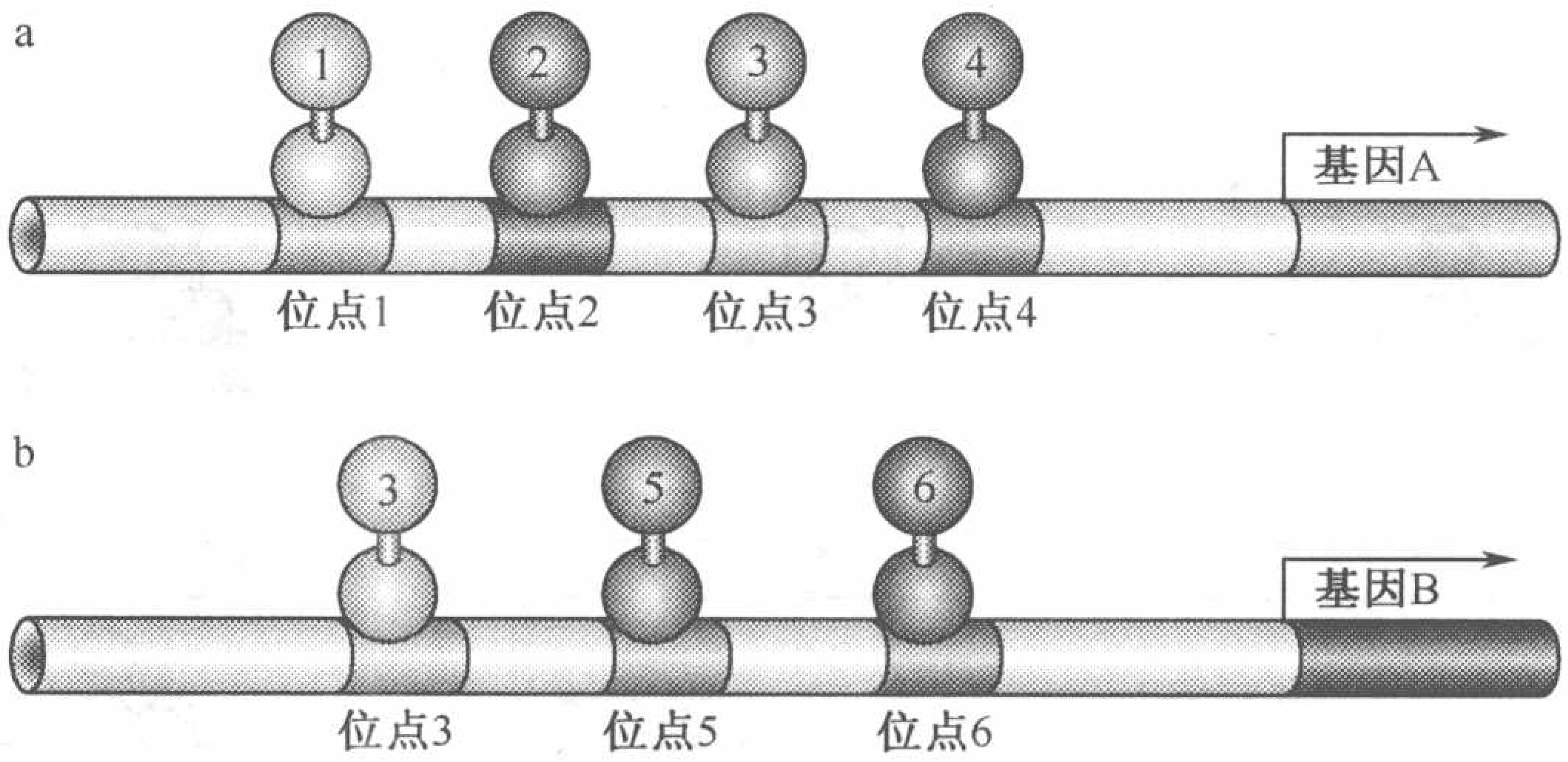


图 17-18　组合控制。这里展示了两个基因，每一个都受到多重信号的控制，基因 A 是 4 种，基因 B 是 3 种。每种信号都是通过一种调控蛋白与基因交流。蛋白 3 与不同的、另外的调节子组合作用于两个基因。

如上所述，多个活化子协同作用。事实上，单个活化子的多个拷贝也能协同作用，表明特定的活化子可以和多个靶体作用。这一点可以解释为什么不同的调节子可以以如此多的组合起作用；每一个调节子都可以利用一系列的靶体，而在一起作用的组合方式并无限制。

以上提到的信号整合的例子——酵母 *HO* 基因和人 β 干扰素基因，所涉及的活化子作为组合控制的例子都调控其他基因的表达。因而，在酵母的例子里，SWI5 涉及多个其他基因的调控。哺乳动物中，NKκB 不仅调控 β 干扰素而且调控许多其他基因的表达包括 B 细胞免疫球蛋白 *κ* 轻链。Jun/ATF 同样与其他调控子作用，调控其他基因的表达。前面讲过某些 DNA 结合蛋白以异二聚体形式与不同的伙伴结合，提供了另外一个层次的组合控制。

酵母交配型基因的组合控制

酵母细胞以 3 种形式存在：两种不同的单倍体交配型 a 和 α，以及当交配型 a 和 α 交配、融合形成的二倍体。两种交配型细胞不同，因为它们表达两套不同的基因：a 特

异基因和 α 特异基因。如下面简述的，它们受到活化子和抑制子的不同组合的调控。

a 细胞和 α 细胞分别编码细胞类型特异的调控子：a 细胞生产调控蛋白 a1，而 α 细胞生产蛋白 α1 和 α2。第四种调控蛋白叫做 Mcm1，也涉及调控交配型特异的基因（以及许多其他基因），它同时存在于这两种细胞中。那么，这些调控子是如何在一起发挥作用从而确保在 a 细胞中 a 细胞特异基因是打开的，而 α 细胞特异基因是关闭的，反之在 α 细胞也是如此。在二倍体细胞中这两套基因都保持关闭。

调控子在 a 细胞特异基因和 α 细胞特异基因启动子上的排列示于图 17-19。

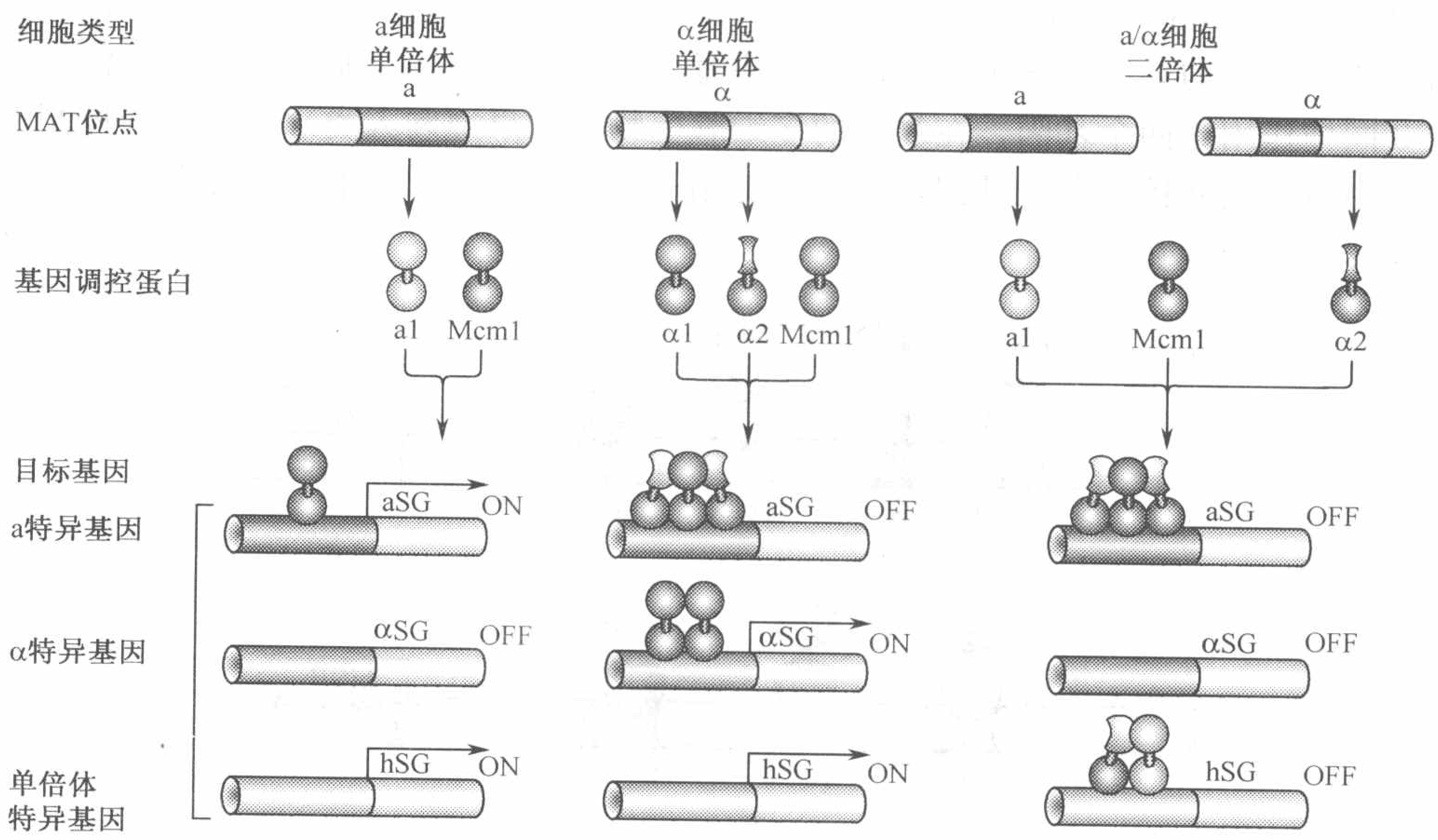


图 17-19 酵母细胞类型特异基因的控制。如在正文中详细讲述的，酵母细胞的 3 种形式［单倍体交配型 a 细胞和 α 细胞以及 a/α 细胞（二倍体）］是由它们所表达的特定的基因来确定的。一种通用的调控子（Mcm1）和 3 种细胞类型特异的调控子（a1、α1、α2）共同调控 3 类靶基因。MAT 位点是基因组中的一个区域，它编码交配类型调控子（第 11 章）。

- a 细胞中，α 细胞特异基因是关闭的，因为没有活化子在那里结合；a 细胞特异基因是打开的，因为 Mcm1 在那里结合并激活这些基因。
- 在 α 细胞中，α 细胞特异基因是打开的，因为 Mcm1 在启动子上游结合并激活它们。在这些基因中，Mcm1 结合在弱结合位点而且只能与蛋白 α1 单体协作发挥作用。这确保了 Mcm1 只在 α 细胞中活化这些基因。α 细胞中 a 细胞特异基因由于 α2 抑制子的存在而保持关闭。这个抑制子以二聚体形式结合并且与 Mcm1 在这些基因中协作发挥作用。α2 抑制子的两种特性保证了 a 细胞特异基因不在这里表达：它覆盖了 Mcm1 的活化区使这个蛋白质无法激活；它同时还有效地抑制这些基因。α2 抑制子的作用机制我们将会在下一节讲到。
- 在二倍体细胞中，a 细胞特异基因和 α 细胞特异基因都关闭。其形成如下：a 细胞特异基因结合 Mcm1 和 α2，与其在 α 细胞中一样，使得这些基因保持关闭。α 细胞特异基因保持关闭，是因为如同在 a 细胞中一样没有活化子的结合。

- 两种单倍体细胞（a 和 α）都表达另外一类基因叫做单倍体特异基因（haploid-specific gene）。这些基因在二倍体细胞中被 α2 关闭，后者在基因的上游与 α1 形成异二聚体与其结合。只有在二倍体中才同时存在这两种调控子。

现在，其他酵母的交配型基因调控的分子细节也已经知道了。在框 17-3 中，我们比较了 a 和 α 特异性基因在酿酒酵母（*S. cerevisiae*）和白色假丝酵母中（*Candida albicans*）是如何调控的。这个比较揭示了基因调控回路可以如何进化，在后面的章节中我们还会回到这个主题。

框 17-3　调控回路的可进化能力

如文中所述（及图 17-19 所示），酿酒酵母的不同交配型以细胞类型特异的方式表达某些基因。在 α 细胞中，α 特异基因表达，a 特异基因不表达；而在 a 细胞中，a 特异基因表达，α 特异基因不表达。我们知道这些程序如何被 *MAT* 位点编码的调控子与泛调控子 Mcm1 共同作用来控制的细节。

白色假丝酵母是酵母的另一个种类，与酿酒酵母有很大的差异。这两个种类大约在 3～9 亿年前有着共同的祖先。如果用保守蛋白的趋异观念来表达的话，这两种酵母的差异就如同鱼和哺乳动物之间的差异。酿酒酵母用于啤酒和面包制作，以及用于实验室的实验，而白色假丝酵母是一种人类病原物。不管怎样，就像酿酒酵母，白色假丝酵母有两种交配型（a 和 α），每种类型以表达不同集的基因为特征（a 特异基因在 a 细胞中表达，α 特异基因在 α 细胞中表达）。此外，表达模式受两个 *MAT* 位点编码的调控子与 Mcm1 的共同作用所控制。

所以，系统的逻辑是一样的，但是控制表达的分子相互作用是不一样的。如我们已经看到的，在酿酒酵母的 a 细胞中，a 特异基因被泛活化子 Mcm1 单独通过那些基因上游的高亲和力结合位点而开启（图 17-19）。因而，没有细胞类型特异调控子的参与。相反，在白色假丝酵母中，a 特异基因的表达确实需要一个细胞类型特异调控子 a2（由这个种类的 *MAT2* 位点编码，酿酒酵母中没有这个基因）。a 特异基因仍然被 Mcm1 激活，但通过一个低亲和力的位点，只有通过 a2 在一个相邻位点协同结合的帮助下才能结合（框 17-3 图 1）。

在酿酒酵母的 α 细胞，a 特异基因被抑制子 α2 的作用而关闭，而在白色假丝酵母中，缺乏 a2 就足以使它们保持关闭。

从白色假丝酵母到酿酒酵母在 a 特异基因调控方式上的进化转变需要 2 个变化：(1) a 特异基因表达变成 a2 非依赖，(2) a 特异基因变成受 α2 负调控。通过对许多酵母种类间的序列比对，可能描绘出下列三个层次的可能次序：DNA 结合位点发生变化、转录调控子丢失、及调控子间蛋白-蛋白相互作用的进化。这些类别中的第一个是所谓的顺式调控变化，后两个是反式调控变化（某种意义上等同于 Jacob 和 Monod 最初在基因调控分析中所定义的；见框 16-2）。

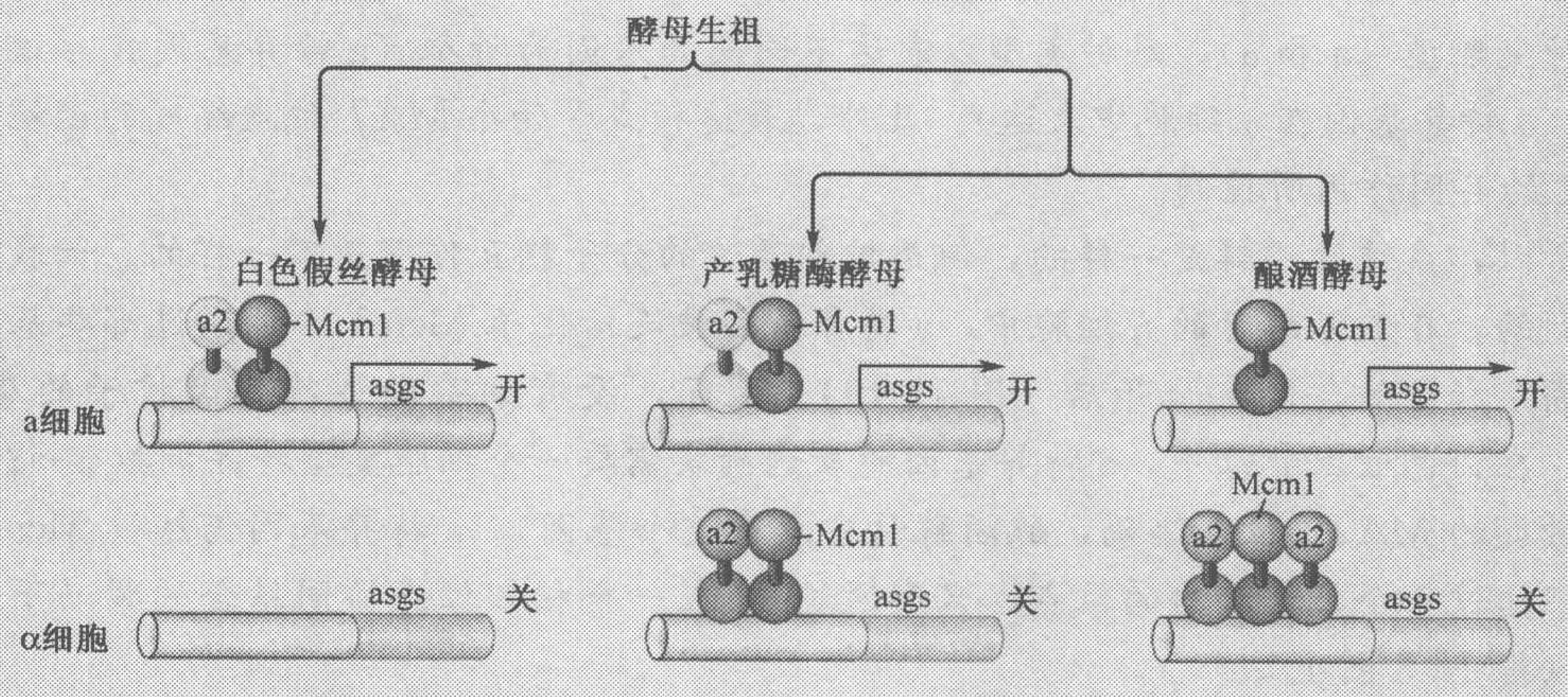
在几个被检测过的种类中，用图来说明其中一个，产乳糖酶酵母（*Kluyveromyces lactis*）（框 17-3 图 1）。这个生物体在进化树上介于酿酒酵母和白色假丝酵母之间，与这层关系相一致，a 特异基因调控模式介于其他两种情况之间。因而，在产乳糖酶酵母

a 细胞中，a2 存在并且结合 Mcm1 来激活 a 特异基因。但在 α 细胞中，α2 也参与这些基因的关闭。

对于 a 特异基因的调控，有人重构了如下的进化转变。在祖先状态，a2 和 Mcm1 活化 a 特异基因。随后，α2 和 McM1 之间的相互作用进化了，还有一个 a2 DNA 结合位点的进化，以及在 a 特异基因启动子中 Mcm1 DNA 结合位点的强度增大。产乳糖酶酵母分异之后，α2-Mcm1 结合特异性增强了，a2 丢失了。

值得注意的是，重构模型中提到的两种调控方式（对白色假丝酵母为正，对酿酒酵母为负）之间转换的中间体有两种形式。这样，在提议的进化路径中所有的中间体都保留了恰当的调控，因为一种新的调控模式在旧的丢失之前被加进来。这引发了另一个可能性：一个特定基因回路的不同版本，如我们已经描述的交配型调控，不是被强烈选择或反对的。每一种都工作得挺好，现代酵母可能继承其中随便哪一个，甚至可能在两者间徘徊。

这里描述的例子涉及了长久以来所有 3 个能改变基因表达方式的变化：顺式调控序列的改变，调控基因出现的改变，调控子间相互作用的改变。这些也正是解释 λ 噬菌体开关（框 16-4）进化所需要的改变。同样的改变也支持了解释大多数动物多样性的调控改变，如在第 19 章所见。反式调控变化的作用在框 19-7 中专门讨论。



框 17-3 图 1 三种酵母中 a 特异基因的调控。（来源：Adapted，with permission，from Rokas A. 2006. *Nature* 443：401-402，Fig. 1. ©Macmillan.）

转录抑制子

在细菌中我们见到许多抑制子通过对与启动子重叠位点的结合从而阻断 RNA 聚合酶的结合。但我们还见到了其他的方式：它们结合在启动子邻近区域，与结合在那里的聚合酶相互作用从而抑制它的转录起始。它们还可以干扰活化子的功能。

除去第一种方式（在细菌中最常见）在真核生物中我们可以见到以上所有其他形式。我们还可以见到另外一种抑制形式，也许是真核生物中最常见的形式。它的作用方

式如下：同活化子一样，抑制子可以募集核小体修饰酶，但在这里这个酶的作用与活化子募集的相反，它们使染色质更紧凑或者去除能够被转录机器识别的基团。例如，组蛋白脱乙酰酶（histone deacetylase）通过从组蛋白尾部去除乙酰基来抑制转录；我们知道乙酰基的存在促进转录。其他的酶在组蛋白的尾部加上甲基，这个反应常常抑制转录。尽管在某些情况下，它与活跃转录的基因关联（第 7 章）。组蛋白（和 DNA）修饰也成为叫做“沉默”的转录抑制的基础，在本章的后面还会详细讲到。

这些各种类型的抑制示于图 17-20。这里我们只对一个例子详细讲述，一个叫 Mig1 的抑制子。与 Gal4 一样，它控制酵母 *GAL* 基因的表达。

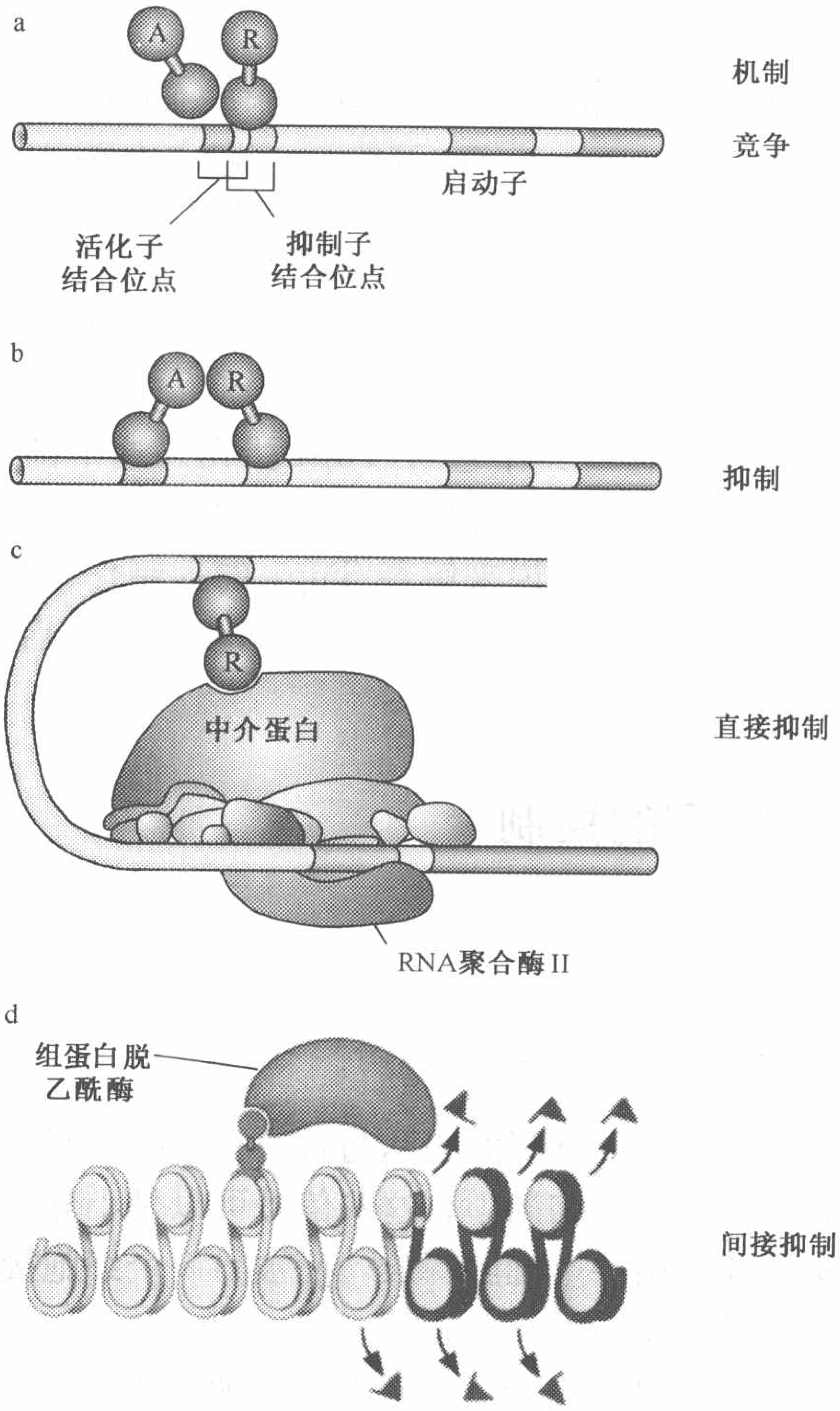


图 17-20 真核生物抑制子的作用方式。真核生物的转录可以受到多种方式的抑制，这包括图中展示的 4 种机制。a. 通过与活化子结合重叠位点的结合，抑制子抑制活化子对基因的结合因而阻断基因的活化。这个模式的变换方式包括：抑制子可以是同一活化子的衍生部分，后者缺少活化区；在另一种模式中，以二聚体形式与 DNA 结合的活化子可以受到其保留了聚合区但缺乏 DNA 结合区的衍生物的抑制，这种衍生物与活化子形成无活性的异二聚体。b. 抑制子与活化子旁边的位点结合，并与活化子相互作用，位阻它的活化区。c. 抑制子与基因上游的位点结合，通过与转录机器的特殊方式作用，抑制转录起始。d. 通过募集组蛋白修饰酶改变核小体，从而抑制转录。（如此例的脱乙酰化，某些情况下也发生甲基化，甚至对启动子的重塑）

图 17-21 展示了我们前面见到的 *GAL* 基因（图 17-3）。但是这里在 Gal4 和启动子之间多了一个位点，这就是在葡萄糖存在的条件下 Mig1 结合并关闭 *GAL* 基因的位点。这样，同 *E. coli* 一样，酵母细胞只有在首选的能量源葡萄糖不存在时才生产代谢半乳糖所需要的酶。那么，Mig1 又是怎样关闭 *GAL* 基因的呢？

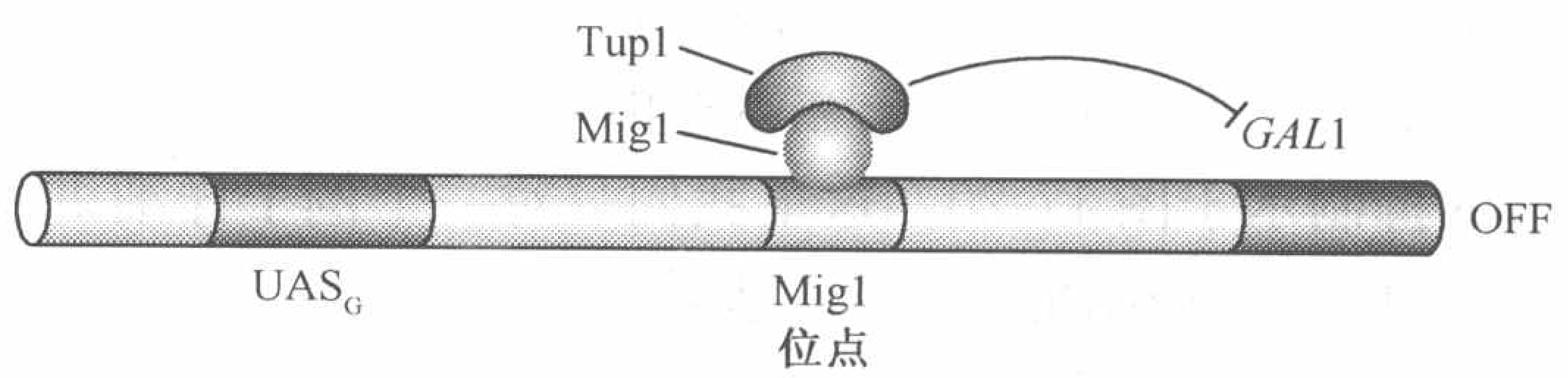


图 17-21 酵母 *GAL1* 基因的抑制。在葡萄糖存在时，Mig1 结合到 UGS_G 和 *GAL1* 启动子之间的一个位点。通过募集 Tup1 抑制复合体，Mig1 抑制 *GAL1* 基因的表达。很可能是对局部核小体的去乙酰化产生了转录抑制（Tup1 募集脱乙酰酶），也可能是直接接触和抑制转录机器。在本图未展示的一个实验中，如果 Tup1 与一个 DNA 结合域融合，而此结合域的位点又位于基因的上游，该基因的表达就会被抑制。

Mig1 募集一个“抑制复合体”其中包含 Tup1 蛋白。这个抑制复合体被其他许多抑制转录的酵母 DNA 结合蛋白所募集，这包括前述的控制交配型的 α2 蛋白。Tup1 在哺乳动物细胞也有对应物。有两种假说来解释 Tup1 的抑制作用：其一，Tup1 募集组蛋白脱乙酰酶，后者使邻近的核小体脱乙酰化；其二，Tup1 在启动子部位与转录机器直接作用抑制转录起始。

信号转导与转录调控子的控制

信号通常通过信号转导通路传达至转录调控子

我们已经知道，某一个特定的基因的表达与否通常取决于其环境中的各种信号。这些信号以多种形式存在。例如，在细菌中，这些信号可以是小分子物质，如糖类。这些信号也可能是蛋白质类物质，由一个细胞释放，而被另一个细胞接受。这在多细胞生物的发育过程中尤为常见（见第 19 章）。

信号被细胞检测到以及信号效应传达到基因有多种方式。在细菌中，信号通过诱导调控子的变构改变来控制调控子的活性。这种效应通常是直接的：小分子信号，如糖类，进入细胞，直接与转录调控子结合。但我们知道一个信号效应是非直接的例子（活化子 NtrC 的控制）。在这个例子中，信号（低氨水平）诱导一种激酶，该激酶磷酸化 NtrC。此类间接信号作用所代表的是信号转导通路（signal transduction pathway）。

“信号”一词指起始配体自身，如糖或蛋白质。这正如我们先前所定义的，“信号”一词也可以用“信息”，因为信号随着配体被识别，传递给调控子，后者直接调控基因，即信号沿着信号转导通路传递。对于简单的细菌例子，“信号”和“信息”没有差别。

但是，一旦涉及信号转导通路，这两者就不同了。在真核细胞中，我们将看到（特别是第 19 章）大多数信号是沿着信号转导通路将信息传递至基因的，这些信号转导通路有时是非常精密的。在此部分，首先介绍一两个真核细胞中信号沿着信号转导通路进行传递的实例。然后，我们更加全面地了解从上述信号转导通路来的信号是怎样控制转录调控子自身的。

在信号转导通路中，起始配体通常由特异的细胞表面受体（cell surface receptor）检测到：配体与受体细胞外的某个区域结合，这种结合的信息被转递至该受体的细胞内区域。从此处开始，信号被传递到相应的转录调控子，通常是通过一系列的激酶级联反应。配体与受体的细胞外区域结合是怎样通讯到细胞内区域？这可以通过受体的变构改变来实现，即配体的结合可以改变细胞内区域的构型（及活性）。此外，配体也可以只是使两条或多条的受体链聚集，使这些受体的细胞内区域间通过交互作用而互相活化。

图 17-22 所示的是信号转导通路的两个例子。首先是一个相对简单的案例——STAT（signal transdncer and activator of transcription）通路（图 17-22a）。在这个例子中，激酶与受体的细胞内区域结合。当受体被其配体（细胞因子）活化，导致两条受体链聚集，引发每条链中的激酶活化，磷酸化对应受体细胞内区域的一段特殊序列。此磷酸化位点随后被一个特殊的 STAT 蛋白识别。而 STAT 蛋白一旦与上述磷酸化位点结合后，其自身也被磷酸化。STAT 蛋白磷酸化后，STAT 蛋白二聚化，向细胞核中转移，并与 DNA 结合。

另外一个例子更为精细（图 17-22b）：控制活化子如 Jun 的 MAP 激酶（MAPK）通路。Jun 是我们前面（图 17-16）提到的 β 干扰素增强子上作用的活化子。在这个例子中，被活化的受体诱导一系列的信号事件，最后导致一种 MAPK 活化，后者进一步磷酸化 Jun（及其他转录调控子）。信息沿着信号转导通路传递，最普遍的方式是通过磷酸化。此外，也会使用蛋白酶解、去磷酸化、以及其他修饰作用。

信号通过多种方式控制真核细胞转录调控子

信号被直接或间接通讯到一个转录调控子后，是如何调控这个调控子的活性呢？在细菌中，控制转录调控子的变构改变通常可以影响该转录调控子与 DNA 结合的能力。这在信号配体自身直接作用于转录调控子的情况以及信号配体经信号转导通路传递到转录调控子的情况中都是如此。因此，乳糖抑制子只有在缺乏别乳糖时才能与 DNA 结合，NtrC 的磷酸化引发变构改变，控制活化子与 DNA 结合。

在真核生物中，通常转录调控子并不在 DNA 结合水平被调控（虽然也有例外），而是通过下述两种基本方式之一被调控。

活化区的暴露。可以通过与 DNA 结合的活化子的构象改变，使先前被掩蔽的活化区暴露出来。或者通过释放掩蔽蛋白达到此目的。在这之前该掩蔽蛋白结合并隐蔽活化区域。上述两种情况所必需的构象改变可以通过（转录调控子）直接与配体结合，或者通过配体依赖的磷酸化来引发。

Gal4 由一个掩蔽蛋白所控制。在缺乏半乳糖时，Gal4 与其 *GAL1* 基因上游的结合位点结合。但是并不活化该基因，因为另外一个蛋白质 Gal80 与 Gal4 结合，闭塞了后

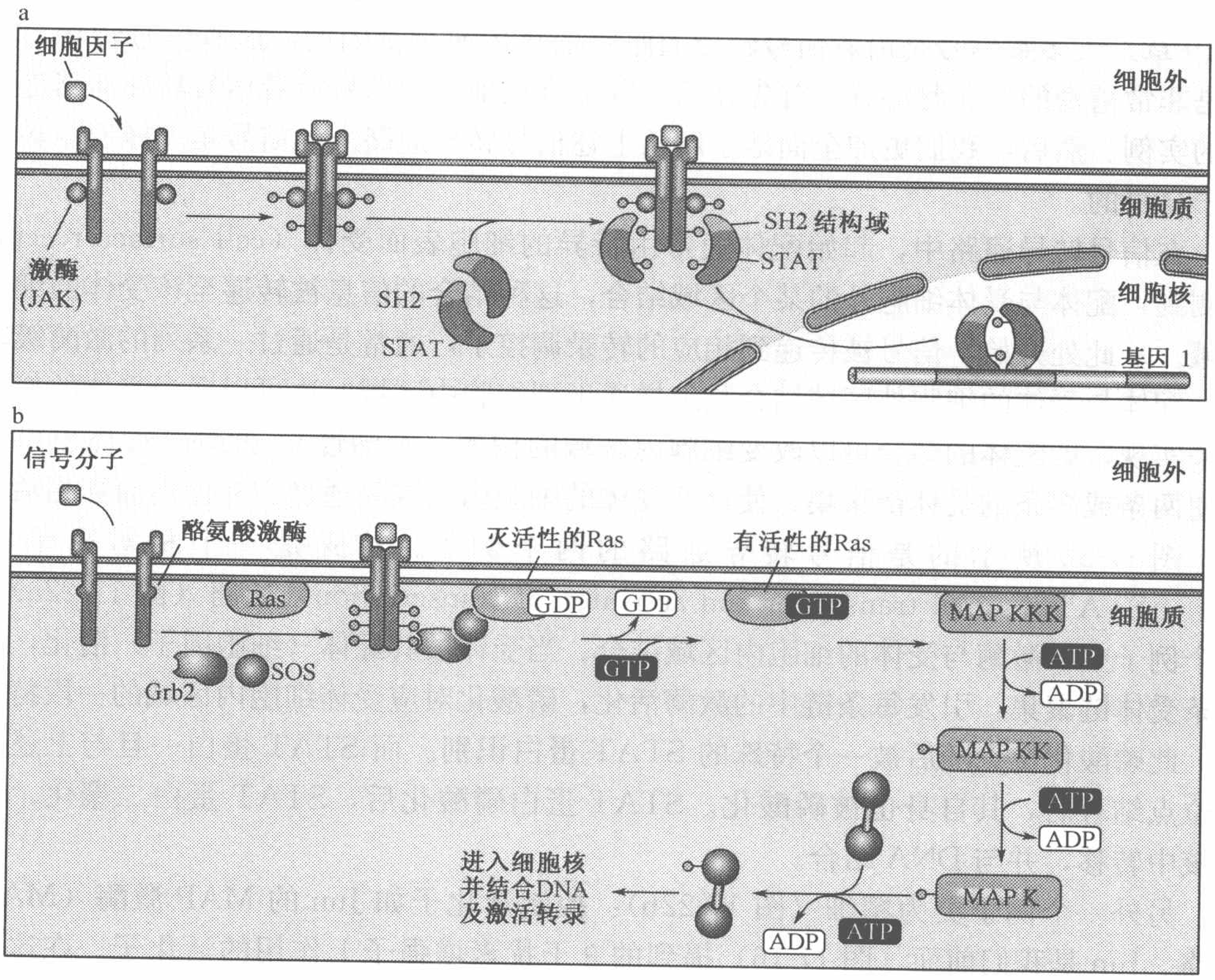


图 17-22 哺乳动物细胞的两条信号转导通路。这里显示的是 STAT 和 Ras 通路。a. 图中显示的是一种细胞因子与其受体结合，从而使两条受体链靠近。每条受体链的胞内区都有一种称为 JAK 的激酶与之附着。两条受体链的会聚（可能伴有由细胞因子结合所引发的构象改变）导致两条受体链的磷酸化，这种磷酸化是由 JAK 激酶所催化的（JAK 激酶也相互磷酸化，激发其激酶活性）。受体链的磷酸化位点随后被细胞质中的叫做 STAT 的蛋白识别。每个 STAT 分子含有一个所谓的 SH2 域，这些结构域见于参与信号转导的很多蛋白中。这些 SH2 域识别在某些序列环境中的磷酸化 Tyr 残基，而这正是该信号转导通路特异性的基础。也就是说，被募集到某一特定受体上的特殊的 STAT 将决定随后被活化的基因。一旦被募集到受体上，该 STAT 自身也被 JAK 激酶磷酸化。这种效应允许两个 STAT 蛋白形成一个二聚体（每个 STAT 上的 SH2 域识别对应 STAT 分子上的磷酸化位点）。二聚体随后转移至细胞核内，在该处，它与 DNA 上特异的位点结合（不同的 STAT 有不同的位点），并且激活邻近基因的转录。b. 引导下游 MAP 激酶通路的 Ras 通路。某一生长因子（如 EGF）与其受体结合，使受体链会聚，正如 STAT 的例子，随后彼此磷酸化。这种作用募集一种衔接蛋白质，叫做 Grb2，该蛋白质有一个 SH2 域，可以识别活化的受体中的磷酸化的 Tyr 残基。Grb2 的另一端与 SOS 结合，后者是一种鸟嘌呤核苷酸交换因子（Ras GEF）。SOS 随后又结合 Ras 蛋白，后者附着在细胞膜内侧。Ras 蛋白是一种小 GTP 酶，后者在与 GTP 结合时采取一种构象，而与 GDP 结合时又是另一种构象；与 SOS 的相互作用引发 GTP 代替 GDP 与 Ras 结合，Ras 因而发生构象改变。在这种新的构象下，Ras 激活所谓 MAPK 级联的最高一级的激酶。该通路中的第一个激酶叫做 MAPK 激酶激酶（MAPKKK）（Raf）；一旦被 Ras 激活后，Raf 磷酸化下一级激酶（一种 MAPK 激酶 MAPKK，叫做 Mek）的丝氨酸和苏氨酸残基。这种作用激活了 Mek，后者进而磷酸化 MAPK（Erk）。该 MAPK 随后磷酸化一系列底物，包括转录活化子（如 Jun），调控多种特异基因的转录，包括 β 干扰素（图 17-16）。

者的活化区。半乳糖触发 Gal80 的释放以及基因的活化（图 17-23）。

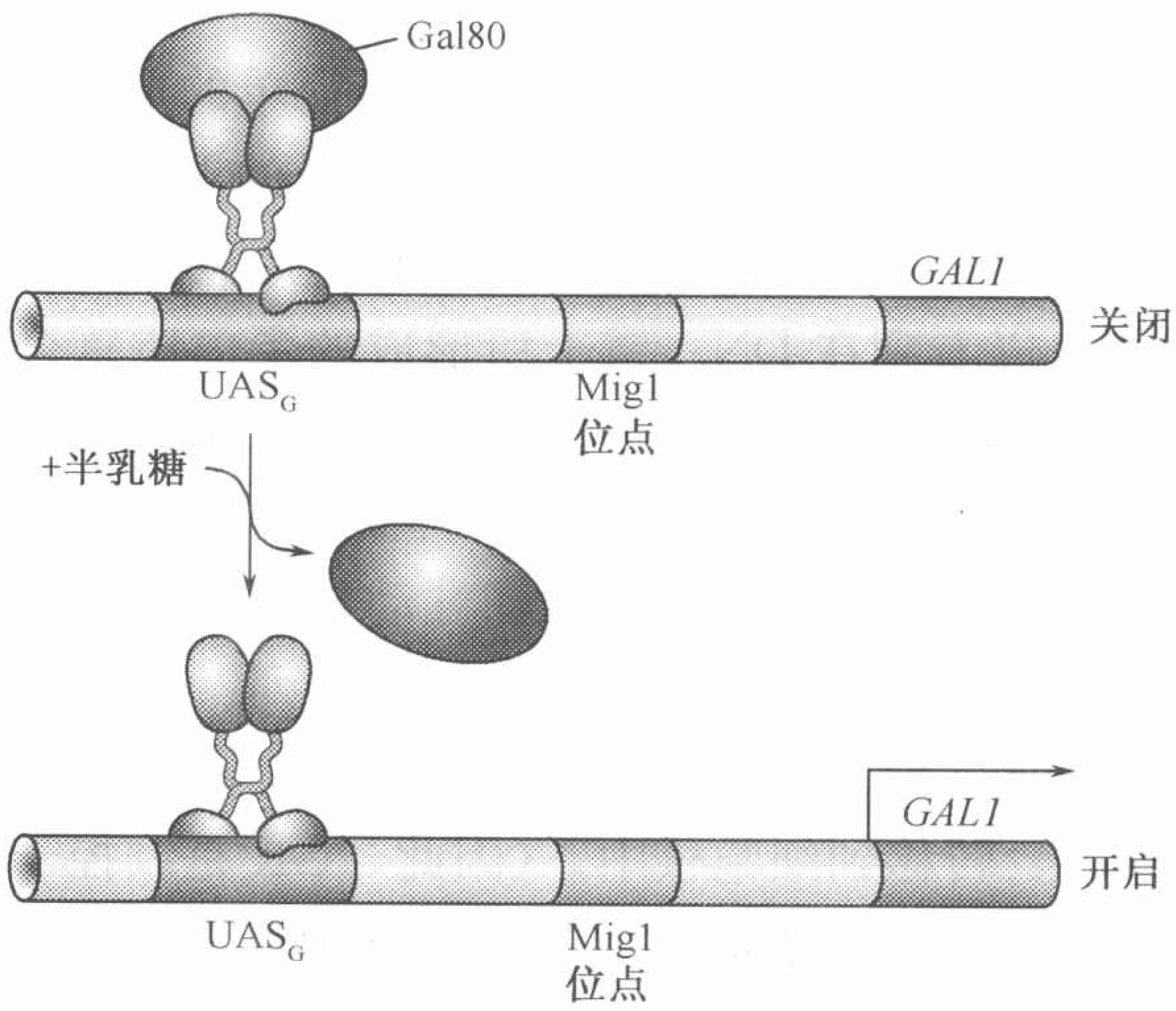


图 17-23 酵母活化子 Gal4 是由 Gal80 蛋白调节的。 Gal4 仅在半乳糖存在时才具有活性，即使在半乳糖缺乏时，发现 Gal4 与其位于 *GAL1* 基因上游的结合位点结合。但是在这种情况下，它并不能活化该基因，因为此时活化区域与另外一个叫做 Gal80 的蛋白质结合。在半乳糖存在时，Gal80 发生构象改变，活化区域因而暴露，*GAL1* 从而被激活。本图显示，在半乳糖存在时，Gal80 与 Gal4 解离。实际上可能是位置的改变以及结合力的减弱，而并非完全脱离。如图显示，Mig1 没有与其位点结合，因为不存在葡萄糖（图 17-21）。

在许多情况下，掩蔽蛋白不仅阻断活化区，其自身作为（或者募集）脱乙酰酶，所以可以主动抑制基因。一个例子是哺乳动物活化子 E2F。E2F 与其目标基因的上游位点结合，无论是否在激活该基因。另外一个蛋白质——抑制子 Rb（retinoblastoma protein，成视网膜细胞瘤蛋白）通过与 E2F 结合而控制 E2F 的活性。这种结合阻断了活化，并且募集了一种脱乙酰酶来抑制目标基因。Rb 磷酸化导致其从 E2F 上释放，因此也就导致了目标基因的活化。E2F 控制某些基因的表达，而这些基因是哺乳动物细胞经历细胞周期的 S 期所必需的（第 7 章）。因此，Rb 的磷酸化控制着这些细胞的增殖。影响这条通路的突变通常都与不受控制的细胞增殖和癌症相关。

运输出入细胞核。当处于非活化状态时，许多活化子和抑制子都被限制在细胞质中。具有信号功能的配体使这些活化子和抑制子转移到细胞核内，它们在细胞核内起作用。这一过程可以有许多不同的机制。因此，调控子可以通过以下方式被限制在细胞质中：与抑制性蛋白相互作用，或者与细胞膜相互作用，或者以某种构象存在，在此种构象中其向细胞核内运输所必需的信号被掩盖了。

对信号的应答所导致的释放和向细胞核内运输可以通过抑制子或连接区域的蛋白质水解以及异构效应来介导。在第 19 章，当我们考虑果蝇胚胎的背腹轴形成时，我们会看到这样的例子。在该例子中，Cactus 是抑制蛋白，在细胞质中与转录调控子 Dorsal 结合。在对某一特异信号应答时，Cactus 被磷酸化后降解，Dorsal 得以进入细胞核行使其功能(图 19-13)。

活化子和抑制子有时以片段形式存在

总的来说，我们已经讨论了活化子和抑制子的最简单的形式，虽然我们已经间接提到额外的复杂度。例如，活化子可以是几个片段活化子的 DNA 结合区和激活区可以分布在不同的多肽链上。它们在 DNA 链上集合到一起，形成活化子。此外，在考虑调控子受信号调控时，我们可以看到在 DNA 链上所形成的蛋白质复合物；而后者的性质可以决定 DNA 结合蛋白是激活还是抑制邻近的基因表达。例如，在上述例子（E2F/Rb）中，活化子在结合某个蛋白质后，转变成了一个抑制子。还有更精细的例子，如糖皮质激素受体（glucocorticoid receptor，GR）。此哺乳动物蛋白质是激活还是抑制转录，取决于其在某一特定基因上的 DNA 结合位点的性质与排列。

在缺少配体时，GR 通过与 Hsp90 蛋白的相互作用，被限制在细胞质中。一旦有配体与之结合，GR 被释放，转移到细胞核中。（因此，GR 是调控子的活性由细胞核定位所控制的另一个实例。）一旦进入细胞核，GR 结合 GRE 位点。这些结合位点有两种类型，当（调控子）结合到其中一类时，它可以激活转录；当（调控子）结合到另外一类时，它可以抑制转录。

当结合到上述第二类位点时，受体采用了某种构象，允许其与组蛋白脱乙酰酶结合。当与上述第一类位点结合时，受体的构象不允许其与组蛋白脱乙酰酶结合，而是与另外一个叫做 CBP 的分子结合 CBP 是被 β 干扰素增强体结合的同一个共活化子（图 17-16）。与 CBP 的结合导致邻近基因的活化。这部分是由于 CBP 自身是组蛋白乙酰酶的缘故，同时也因为募集构成转录机器的某些成分，就如同我们在 β-干扰素例子中描述的那样。事实上，在激活过程中，CBP 被募集到许多哺乳基因上。

“共抑制子”和“共活化子”通常指任何辅助性蛋白，既非转录机器的一部分，其自身也不是与 DNA 结合的调控子，但是却参与基因的转录调控。CBP 就是这样的例子。此概念也指其他对核小体具有修饰功能的复合物。

组蛋白与 DNA 修饰导致的基因“沉默”

到目前为止，我们已经讨论了活化子和抑制子对基因的调控，它们与基因邻近区域结合，开启或关闭该基因。这种效应是局部的，并且调控蛋白的活动通常由特异的细胞外信号控制。现在我们转向讨论转录沉默（transcriptional silencing）的机制。沉默在本文中（在第 18 章这个词被用于完全不同的情形）指的是一种位置效应。基因因为它所处的位置而沉默，而不是由于对环境中某一特异信号的应答反应。另外，沉默效应在大段的 DNA 序列上可以“扩散”，关闭多个基因，即使是那些离起始事件距离较远的基因。尽管有这些差异，了解沉默效应并不需要全新的理论，而仅仅是我们在本章所遇到的理论的延伸。

最常见的沉默形式与被称为异染色质（heterochromatin）的一种致密形式的染色质有关。异染色质的命名源于其在光学显微镜下的外观：与其他染色质，即常染色质

(euchromatin)相比，它看起来很致密。异染色质通常与染色体的某些特殊区域相关，特别是端粒（在染色体末端发现的结构）和着丝粒。正如在第7章所学到的，端粒和着丝粒通常由重复序列组成，偶尔会有少量的蛋白质编码基因。如果用实验的方法将一个基因转移至该区域，那么该基因通常被关闭。实际上，还有其他区域的染色体也处于异染色质状态，并且这种区域中已发现有基因的存在，如酵母中沉默的交配型基因座。在哺乳动物细胞中，估计约有50%的基因组以异染色质的某种形式存在。

我们已经知道，染色质的密度可以被那些在化学上修饰组蛋白尾部的酶所改变。这种修饰影响DNA的易接近性，并由此影响复制、重组以及转录等过程。

正如我们已经叙述过的，转录过程的激活和抑制通常涉及核小体的修饰，以至于改变基因对转录机器和其他调控蛋白的易接近性。我们也曾经遇到过某些蛋白质可以识别被修饰过的核小体，并且可以特异性地与之结合。异染色质的沉默可以被理解为上述原理的一种延伸。

转录也可以通过DNA甲基化而被沉默。DNA甲基化是由DNA甲基化酶（DNA methylase）所催化。这种形式的沉默在酵母中尚未发现，但在哺乳动物细胞中却很普遍。DNA序列的甲基化能抑制蛋白质结合，包括转录机器，因而阻断基因的表达。但是甲基化也可以通过其他途径抑制表达：有些序列只有在被甲基化后才能被特异的抑制子识别，这种识别随后关闭邻近的基因，通常通过募集组蛋白脱乙酰酶来实现。

在酵母中，沉默由组蛋白的脱乙酰化作用和甲基化作用介导

在酿酒酵母中，端粒、沉默的交配型基因座以及rDNA基因都是"沉默"的区域。我们以端粒为例进行讨论。

现已发现，每条染色体末端1～5kb部分均是折叠、致密的结构，如图17-24所示。来自染色体其他位点的基因转移到该区后，通常都变成沉默状态，特别是当它们在原先的位点就只有微弱表达。相比其余部分的基因组（这部分基因表达更容易），端粒部分的染色质乙酰化程度要轻一些。

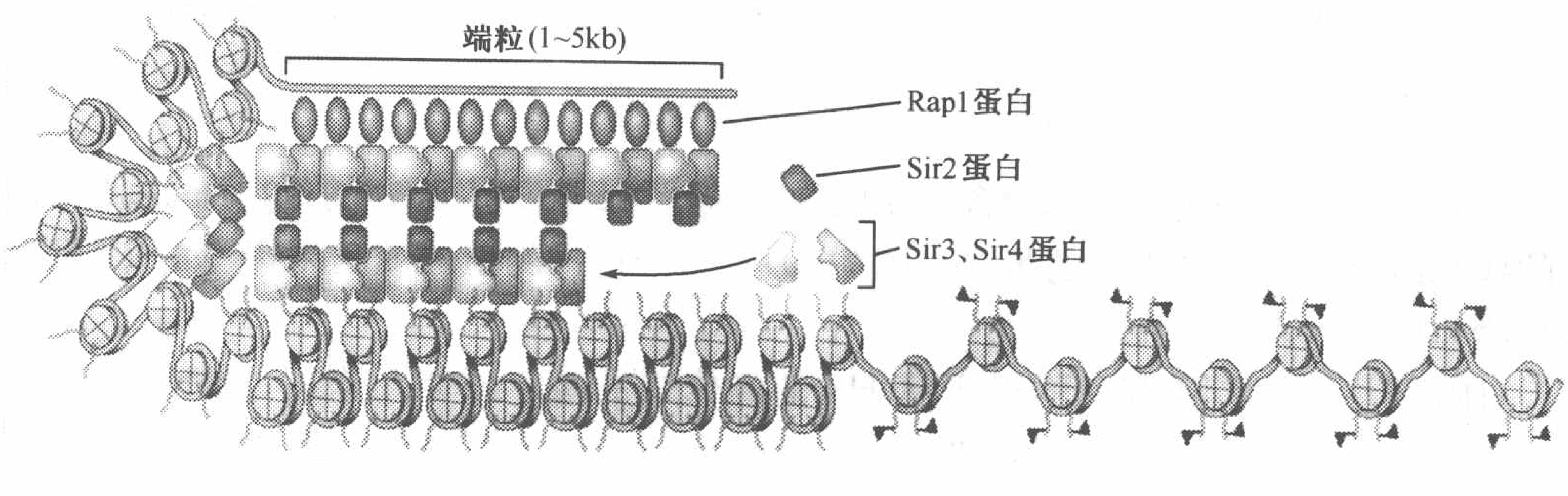


图17-24　酵母端粒中的沉默效应。Rap1募集SIR复合体至端粒。SIR2，该复合体的一种成分，使邻近的核小体脱乙酰化。未被乙酰化的尾部随后与SIR3及SIR4结合，募集更多的SIR复合体，从而允许其中的SIR2作用于更远处的核小体，循环往复。这就解释了由脱乙酰化造成的沉默效应的扩散。（来源：Adapted，with permission from Grunstein M. et al. 1998. Cell 93：325-328. © Elsevier.）

已经分离到沉默效应被消除的突变体，即处于端粒部位的基因以更高的水平表达。这些研究显示了 3 个编码沉默调控蛋白的基因，即 *SIR2*、*SIR3* 和 *SIR4*（*SIR* 代表 silent information regulator）。这 3 个基因所编码的蛋白质形成一个复合物，与沉默状态的染色质结合，其中 Sir2 是一种组蛋白脱乙酰酶。

沉默复合物由识别端粒重复序列的 DNA 结合蛋白募集到端粒。在沉默的交配型基因座，募集也是由一个特异的 DNA 结合蛋白起始的。上述两种情况中，Sirs 的募集激发了组蛋白尾的局部脱乙酰化。胶乙酰的组蛋白随后被沉默复合物直接识别，因此局部脱乙酰化以自我永续的方式沿着染色质轻易地扩散开，产生了一个扩展的致密异染色质区域。

与 Tup1 的抑制（其机制仍未明了）不同，沉默很清楚地是由组蛋白尾的脱乙酰化造成的。Sir2 的丢失完全缓和了沉默，而且组蛋白尾的乙酰化具有类似的效应。整个异染色质结构依赖于 DNA 结合蛋白（Rap1）的持续存在来维持完整性。因此，尽管因 Sir 识别脱乙酰组蛋白而引发脱乙酰化的加剧和扩散，DNA 结合蛋白继续扮演重要的角色。另外，是 DNA 结合蛋白赋予了整个过程的特异性，即订定了沉默复合物在哪里形成。在某些沉默例子里，RNA 分子，而不是蛋白质，提供了这个重要的特异性。在第 18 章，我们将讨论这样一个例子，另一种酵母（粟酒裂殖酵母 Schizosaccharomyces pombe）的 RNAi 是该生物体交配型基因座和着丝粒沉默所必需的。

沉默区域的扩散是如何被节制的？即它是如何被限制到合适的区域（如端粒、沉默的 MAT 基因座等）？前面我们提到绝缘子元件能够阻碍组蛋白修饰的扩散（图 17-12）。另外，其他种类的组蛋白修饰阻碍 Sir2 蛋白的结合，于是终止了扩散。组蛋白 H3 尾的甲基化相信起了这个作用。

组蛋白甲基转移酶把甲基基团加到组蛋白尾上。如我们在第 7 章看到的，这些酶将甲基基团加到蛋白 H3 和 H4 尾上特异的赖氨酸残基上。组蛋白甲基转移酶最近在酿酒酵母中有所描述，被认为帮助某些基因的抑制，如刚看到的，阻碍 Sir2 介导的沉默在其他区域的扩散。但是组蛋白甲基化酶在更高等的真核生物和粟酒裂残酵母中被研究得更详细。在这些生物体中，沉默通常与含有既脱乙酰又甲基化的组蛋白的染色质相关联。于是在这些生物体中，在 H3 尾 Lys-9 的甲基化是与沉默的异染色质相关的一种修饰（表 7-7）。相反，其他位点的甲基化（如在同一尾上的 Lys-4）与表达的提高相关。

果蝇中 HP1 识别甲基化的组蛋白和凝聚染色质

如同组蛋白乙酰化的残基被含有溴区结构域（bromodomain）的蛋白所识别，甲基化的残基结合含溴区结构域的蛋白（图 7-40）。一个这样的蛋白就是果蝇蛋白 HP1，是该生物体中沉默异染色质的一个组分。

HP1 蛋白与含有甲基化的组蛋白 H3 的经修饰的染色质相互作用。这个特定的修饰是由 Su（Var）3-9（一个所谓花斑化（variegation）的抑制子）编码的一个酶产生的。花斑化在某些情况中当一个基因被转移到一个异染色质区域时可见。那个基因不是在所有细胞所有时间都沉默着，而是在沉默和表达状态间明显地随机跳来跳去，在某些细胞中“开启”，而在其他细胞中“关闭”。花斑化在所谓的 *white* 基因尤其明显，该基因负责成蝇眼睛正常的红色色素化。这些基因被称为 *white*（白色）是因为突变表型是

白眼（第22章）。当插入到异染色质，*white*基因的表达变成“花斑化”，产生的眼睛带有花白的红色。*Su*（*Var*）3-9基因的突变抑制了这种花斑化，产生的眼睛有着更统一的红色；在那么多细胞里，*white*基因的表达不再沉默了。Su（Var）3-9蛋白是一个组蛋白H3 Lys-9甲基转移酶。通过目前还不明白的机制，它在组蛋白H3尾上加上甲基基团。这个修饰对HP1蛋白的结合是必不可少的，HP1蛋白随后参与异染色质的凝聚过程。Su（Var）3-9被认为还可以由序列特异的DNA结合蛋白募集到特异的异染色质，因而导致由HP1介导的基因特异的组蛋白甲基化和转录抑制。

我们已经看到单一种类的修饰如何能参与到基因调控中。但是当多种类型的修饰在同一个基因上发生会怎样呢？它们的效应如何相互作用？有人提议复杂模式的修饰按“组蛋白密码”来操作。各种组蛋白修饰间的相互作用和组蛋白密码的概念在框17-4中描述。

框17-4　存在组蛋白密码吗?

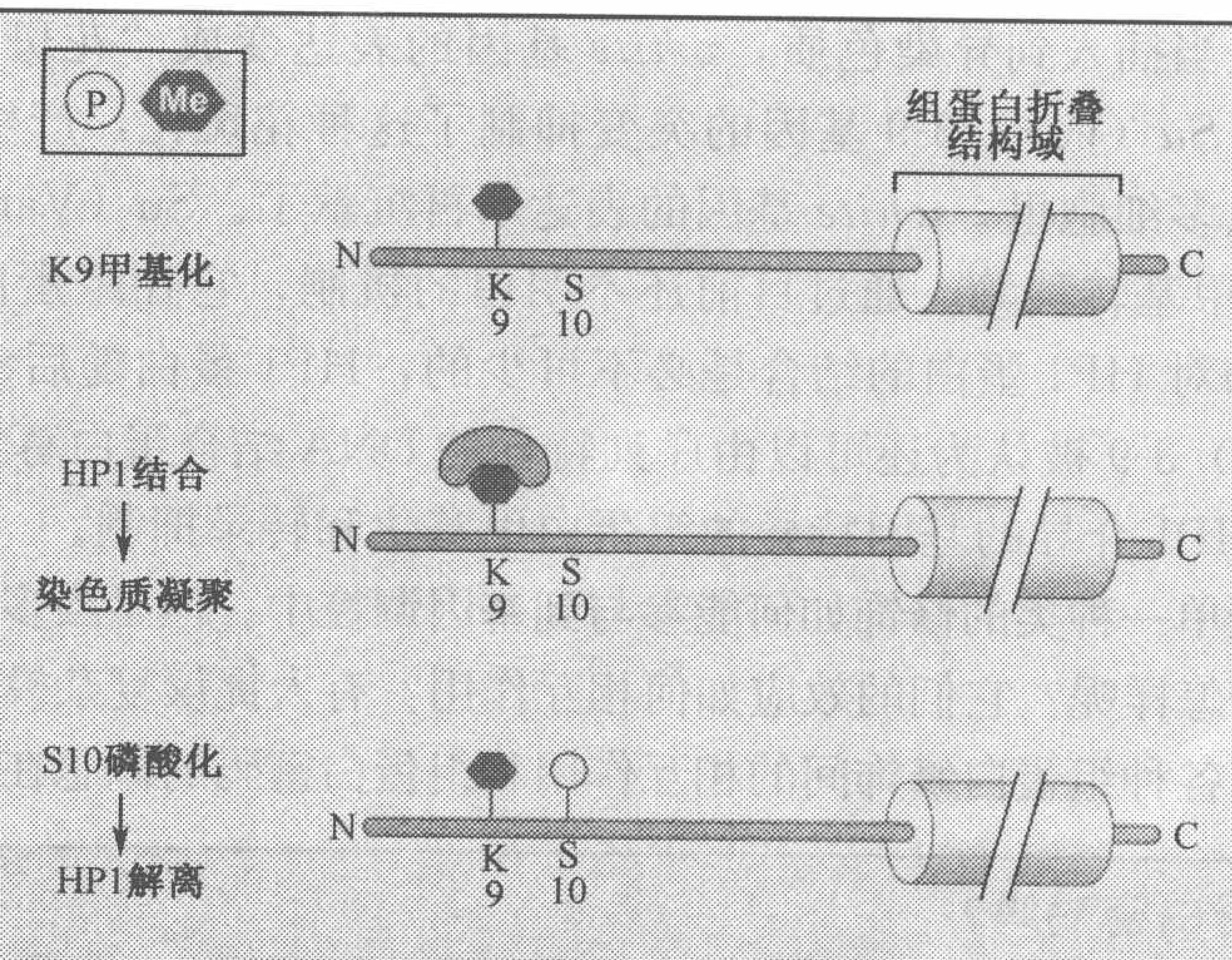
有提议存在组蛋白密码（histone code）。根据这个概念，在一个特定基因组蛋白尾上的不同模式的修饰可以被“阅读”，表明不同的意思（图7-39）。其中的“含义”来自于每个募集不同组蛋白的特定修饰形式；特定的组合形式取决于这些蛋白所携带的识别结构域的数量、类型和组成。

我们已经遇到过在组蛋白上识别特异的乙酰化和甲基化“标记”的蛋白（如，TFⅡD和HP1）。还存在H3和H4尾上磷酸化丝氨酸残基的蛋白，以及结合这些修饰的蛋白。因而在组蛋白尾的几个位置上进行多种修饰是可能的（图7-39）。另外许多携带修饰识别结构域的蛋白本身就是酶，它们进一步修饰组蛋白。我们开始看到，识别和维持修饰模式的过程在理论上是如何实现的。

考察一个简单的例子——在组蛋白H3尾的Lys-9（图7-39）。这个残基的不同修饰状态可以被解读成不同的意思。于是，这个残基的乙酰化与活跃转录的基因相关。这个残基被各种携带溴区结构域的组蛋白乙酰化酶所识别，刺激邻近核小体的乙酰化。当Lys-9没有被修饰时，它与沉默的区域相关（如我们前面在酿酒酵母中所见）。非乙酰化的组蛋白通常募集脱乙酰酶，加强和维持脱乙酰化状态（如我们在酿酒酵母沉默区域的扩散中所见）。最后，在某些生物体中同一个赖氨酸可以被甲基化：在那种情况下，修饰的残基与建立和维持异染色质状态的蛋白（如HP1）结合。

但是修饰的组合会有其独特的含义吗？再次拿果蝇HP1蛋白来作为一个组蛋白的修饰如何明显地影响邻近第二个组蛋白的修饰的例子来说明（框17-4图1）。在细胞分裂中期，HP1蛋白暂时离开有丝分裂染色体，即使它们保留了基本的“标记”，即组蛋白H3 Lys-9的甲基化。失去HP1的结合与旁边H3 Ser-10的磷酸化相关。这个磷酯化由一个细胞周期激酶AuroraB介导。该酶只有在细胞周期的M期变得有活性，因此导致HP1从分裂中期染色体的异染色质上脱离。

HP1被AuroraB激酶解离看上去是将有丝分裂纺锤体搭上着丝粒及随后在胞质分裂中姐妹染色单体的分离所必需的。当这个过程完成后，由于AuroraB激酶活性的消失，Ser-10的磷酸化就消失了，HP1重新与染色体相互作用维持异染色质。与这个模型相一致，删除AuroraB激酶的突变导致异常的分离。



框 17-4 图 1 一个染色质修饰对另一个染色质修饰的影响。图示在组蛋白 H3 上的修饰。Lys-9（K9）的甲基化募集了 HP1，后者随后作用于染色质的凝聚。邻近丝氨酸残基（S10）的磷酸化从甲基化 Lys-9 移开 HP1，而没有移走甲基基团。

尽管看到这些现象，存在特异密码，即在一个特定基因座上复杂的组蛋白修饰产生一个高度特异的信息，仍然具有高度的争议。在一个基因上看到的许多修饰可能仅仅是一个基因在被活化或被抑制的过程的一部分，而不是作为起始信号。即它们是基因正在“开启”或“关闭”的结果，而不是原因。当一个特定基因被表达时，位点特异性 DNA 结合蛋白（或，在某些情况下小 RNA 分子，如我们将在第 18 章中所见）继续是决定异性的最强大的提供者。

在哺乳动物中，DNA 甲基化与沉默状态的基因有关

有些哺乳动物基因由于其邻近 DNA 序列的甲基化而处于沉默状态（我们现在谈论 DNA 甲基化，而非组蛋白甲基化）。事实上，哺乳动物基因组上大段的区域正是通过这种方式被标记的，并且 DNA 甲基化通常出现在异染色质区域。这是因为甲基化的序列常被 DNA 结合蛋白（如 MeCP2）识别。DNA 结合蛋白可以募集组蛋白脱乙酰酶和组蛋白甲基化酶，而这两种酶可以修饰邻近的染色质。因此，DNA 甲基化可以标记随后异染色质形成的位点（图 17-25）。

DNA 甲基化在一种被称为印记（imprinting）的现象中处于核心地位。在一个二倍体细胞中，多数基因只有两份拷贝。其中一条染色体上的一份拷贝来自父方，另外一条对应染色体上的拷贝来自母方。在多数情况中，两个等位基因表达水平具有可比性。这并不奇怪：它们携带同样的调控序列，而且处于同样的调控子中；它们在两条非常相似的染色体中也位于相应的区域。但在少数情况下，只有其中一份拷贝的基因表达，而另一份拷贝处于沉默状态。

两个研究的比较透彻的例子是人 *H19* 基因和胰岛素类似生长因子 2（*Igf 2*）基因（图 17-26）。这两个基因位于人类第 11 号染色体上相互邻近的位置。在某一特定的细胞

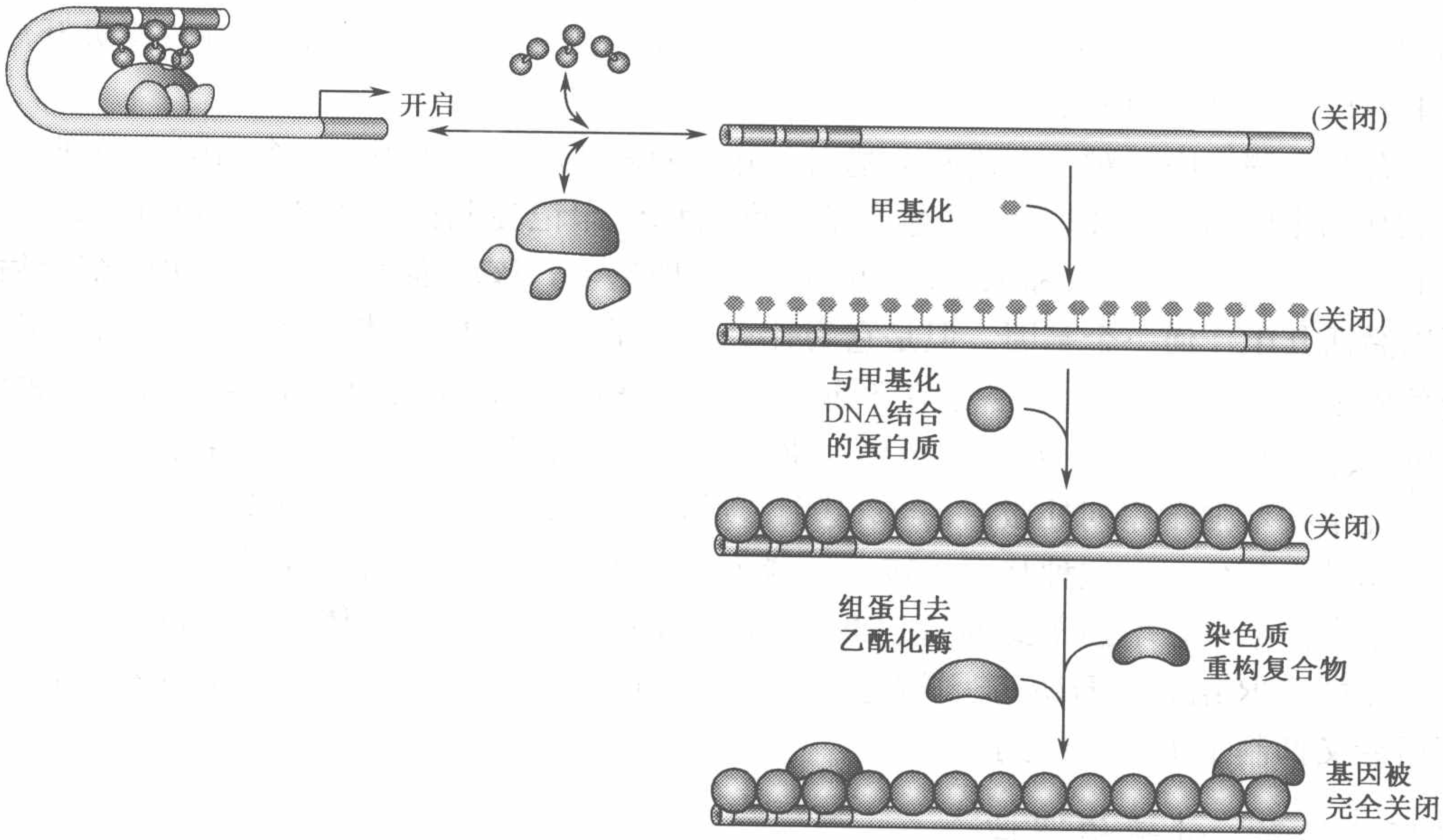


图 17-25 DNA 甲基化与组蛋白修饰关闭基因的表达。在未修饰状态，活化子和转录机器存在时，哺乳动物基因可以很容易在表达状态和非表达状态之间转换，如顶行所显示。在这种状态下，表达从来不是被完全关闭，而是具有渗透性。通常这并不是最好的状态，有时基因必须被完全关闭，偶尔被永久性的关闭。这是通过 DNA 的甲基化和局部核小体的修饰来实现的。因此，当某基因不被表达时，DNA 甲基转移酶（一种甲基化酶）可以接近并甲基化启动子序列、基因自身和上游活化子结合位点序列内部的胞嘧啶。甲基基团被加到胞嘧啶环的 5′ 位置上，生成 5′-甲基胞嘧啶（第 6 章）。有的情况下，这种修饰自身能破坏转录机器以及活化子与基因的结合。但是它也可以增强其他可以识别含甲基胞嘧啶的 DNA 的蛋白（如 MeCP2）的结合。这些蛋白质进而又可以募集某些可以重塑和修饰局部核小体的复合物，从而完全关闭基因的表达。

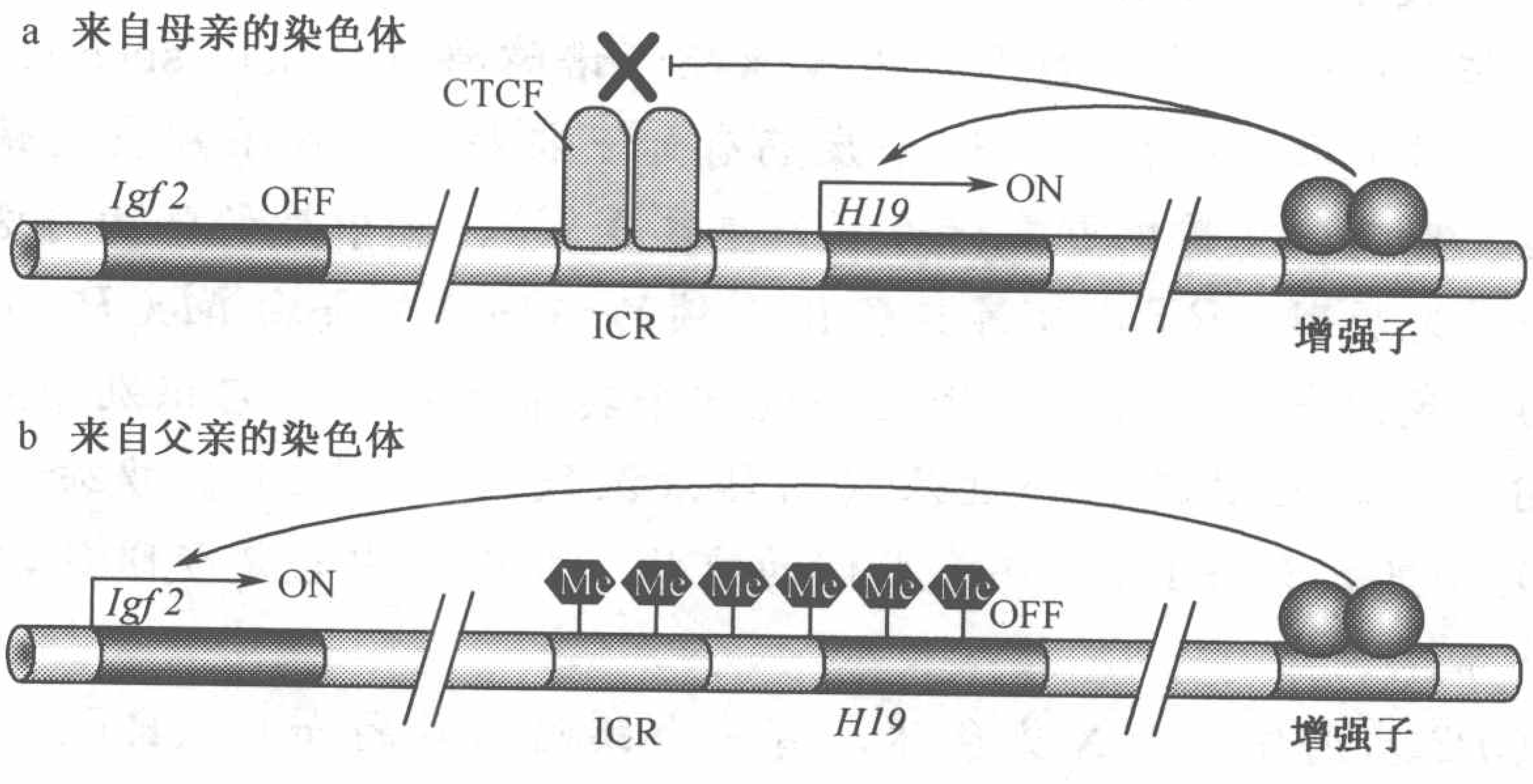


图 17-26 印记。这里显示的是由印记所控制的两个基因——哺乳动物 *Igf2* 基因和 *H19* 基因。如文中所述，在某一个特定的细胞中，只有母本染色体上的 *H19* 基因表达，而 *Igf2* 只从父本染色体上的基因表达。绝缘子元件的甲基化状态决定了 ICR 结合蛋白（CTCF）是否可以结合进而阻断 *H19* 基因被下游的增强子所激活。

中，一份拷贝的 *H19* 基因（位于母方来源的染色体上）表达，而另一份拷贝的基因（位于来自父方的染色体上）则被关闭。*Igf 2* 基因则正相反——来自父方的拷贝开启，而来自母方的拷贝则关闭。

有两个调控序列对于上述两个基因差异性的表达起关键作用：一个增强子序列（位于 *H19* 基因的下游）以及一个绝缘子序列（称做印记控制区 [ICR，Imprinting Control region]，位于 *H19* 和 *Igf2* 基因中间)。理论上，该增强子（当与活化子结合后）可以激活其中任何一个基因。但是它为什么仅仅活化母本染色体上的 *H19* 基因和父本染色体上的 *Igf2* 基因呢？答案在于 ICR 及其甲基化状态。增强子不能激活母本染色体上的 *Igf2* 基因，是因为在该染色体上 ICR 与一种叫做 CTCF 的蛋白质结合。CTCF 阻断活化子在增强子序列处激活 *Igf2* 基因。反之，在父本染色体上，ICR 元件和 *H19* 启动子均被甲基化。在这种状态，转录机器不能与 *H19* 启动子结合，而且 CTCF 不能与 ICR 结合。结果，增强子此时可以激活 *Igf2* 基因。在父本染色体上，通过 MeCP2 与甲基化的 ICR 结合，*H19* 基因被进一步抑制。同前，这种作用可以募集脱乙酰化酶，而后者又抑制了 *H19* 启动子。

基因表达的 epi 遗传调控

框 17-5 描述了抑制丧失导致人类疾病的 2 个例子。一个涉及 MeCP2，另一个涉及印记缺损。

框 17-5 转录抑制与人类疾病

有几个人类疾病由特异基因的去抑制而引起。这里，我们考察 2 个疾病。每一个都由一个基因的抑制丧失而引起，基因抑制的机制在文中有描述。第一个是 Rett 综合征，由抑制蛋白 MeCP2 的缺失而引起。第二个是 Beckwith-Weidemann 综合征，由在 *Igf2* 基因的 ICR 上缺失了 CTCF 结合位点而引起。

Rett 综合征（RTT）是一个严重的孤独症频谱障碍（autism spectrum disorder），每 10，000 个女孩中有一个得病。这个疾病有如下特征：儿童早期丧失语言和运动技巧、头小畸型、癫痫、反常性重复行为（如重复手写）、和中度的强力呼吸。它是零星的智力迟钝的常见原因。RTT 由 X 染色体连锁的编码抑制蛋白 *MeCP2* 的基因上的一个突变引起的。我们先前在文中已经遇到过这个转录调控子。它识别甲基化的 DNA，并通过募集组蛋白脱乙酰酶使邻近基因的转录沉默（图 17-26）。携带一个中断了的 *MeCP2* 基因的小鼠具有与 RTT 病人类似的症状。在小鼠中它是受限的，MeCP2 的缺失只局限在脑部。

因为 *MeCP2* 基因发现于 X 染色体，有一个缺陷拷贝的女孩（RTT 病人）是一个嵌合体。在那些携带该基因突变拷贝（同位基因）的 X 染色体失活的细胞里，野生型 MeCP2 被制造了。但在那些野生型拷贝（同位基因）的基因位于失活的 X 染色体上的细胞里，就没有 MeCP2 被制造出来了。在（单一）X 染色体上携带突变基因的男孩在所有细胞里都缺乏 MeCP2，通常在一到两年内死于呼吸衰竭。

RTT 被认为是一个神经发育疾病，而非神经退化疾病，因为病人（及基因敲除小

鼠）显现异常的神经元形态，而不是神经死亡。即便如此，MeCP2 缺乏被认为可能在发育的某个特殊点上是重要的，此后即使重塑它的功能也无法逆转这个表型。但是最近构建了一个小鼠，其 MeCP2 的表达能够被操纵，使得该小鼠在没有 MeCP2 表达的情况下长到成年，早于该调控子表达被开启。引人注目的是，在成年制造 MeCP2 足以逆转早期发育过程中缺乏所带来的效应。这个令人兴奋的发现使在人类中进行治疗干预变得更可行，即使仍然很困难。

MeCP2 与 RTT 症状之间的联系没有完全被搞明白。如我们已经看到的，MeCP2 是基因表达的一个抑制子，它的目标基因之一编码脑源性神经营养因子（brain-derived neurotrophic factor，BDNF）。这个蛋白是一个生长因子，在脑部发育及与学习和记忆相关联的突触改变中扮演角色。最近发现神经活性导致 MeCP2 的磷酸化，该修饰导致抑制子从 DNA 上解离，大概使得它的目标基因得以表达。中断这些活性（如，BDNF 不适当地表达）作为对至少某些 RTT 的认知症状的解释具有明显的吸引力。

对 MeCP2 和 BDNF 间的联系及这些蛋白质与疾病之间的联系的研究正在进行，对由 MeCP2 调控的其他可能的相关基因的研究也在进行。从认知障碍到异常步态的一系列广泛症状提示许多基因的表达异常可能是疾病的完全发作所必需的。

Beckwith-Wiedemann 综合征（BWS）是一个发育病症，在每 15，000 人中有一人得病。这个疾病有如下特征：生长过度（有这个疾病的儿童早产且大于正常胎儿）、对各种儿童癌症（包括肾母细胞瘤）易感性增加。其症状还与在染色体 11p15.5 上的印记基因的中断表达有关联，其中包括在文中讨论过的胰岛素类似生长因子 2 基因（*Igf2*）（图 17-26）。

如我们已经描述过的，*Igf2* 基因通常只由一个同位基因（父本）表达，因为另一个被印记了。同时，*H19* 基因只从母本同位基因表达。许多 BWS 的病例与 *Igf2* 的双同位基因表达及 *H19* 基因的不表达相关联，这是两个染色体 ICR 都被甲基化的结果。IGF2 是一个致命的生长因子，H19 相信是一个参与肿瘤抑制的调控 RNA（第 18 章），因此该病的表型（生长过度和肿瘤敏感）可以理解。

与 RTT 一样，BWS 的症状在适当操纵的小鼠里被模拟：过表达 *Igf2* 产生的结果是，小鼠普遍生长过度并且出现特异肿瘤。

基因的表观遗传调控

基因表达的模式有时必须得到继承。在发育中一个细胞释放的信号导致邻近细胞开启特异的基因。这些基因也许不得不在这些细胞的许多代内都保持开启，即使诱导它们开启的信号只瞬间存在。在同时缺乏突变和起始信号的情况下，基因表达模式的继承称为表观遗传（epigenetic）调控。

与之相反的是我们已经讨论过的某些基因调控例子。如果一个基因由一个活化子控制，而且那个活化子是在一个特定的信号存在时才有活性，那么，只要信号还存在，基

因就会保持开启。事实上，在正常的条件，E. Coli 的 *lac* 基因只在有乳糖存在而葡萄糖缺乏的时候才表达。同样的，酵母的 *GAL* 基因只在葡萄糖缺乏而半乳糖存在时才表达，人类 β 干扰素只在细胞被病毒感染刺激时才制造。

基因表达的某些状态跨越细胞分裂得以继承，即使起始信号不再存在

我们已经遇到过基因调控可以被表观遗传继承的例子。考察一个噬菌体 λ 溶源菌的维持（第 16 章）。在溶源菌中，噬菌体在细菌宿主细胞内处于休眠状态。这个状态与基因表达的一个特异模式相关，尤其是与 λ 抑制子蛋白的持续表达相关（图 16-26）。

溶源性基因表达是在一个感染细胞中应答贫瘠的生长条件而建立起来的。然而，一旦建立，溶源性状态就稳定地维持，尽管生长条件有所改善：将溶源菌转移到富含营养的生长培养基并不会导致诱导作用。事实上，诱导基本上从不发生，直到接收到一个合适的诱导信号（如 UV 光照）。

溶源性状态跨越细胞分裂得以维持便是表观遗传调控的一个例子。这个表观遗传控制源自于一个抑制子合成的两步策略。第一步，抑制子合成通过活化子 CⅡ（其对生长条件敏感）对抑制子（CⅠ）的激活而建立。第二步，抑制子合成通过自动调控得以维持：抑制子激活自身基因的表达（图 16-30）。通过这条途径，当溶源性细胞分裂时，每个子细胞继承了一个拷贝的处于休眠状态的噬菌体基因组和一些抑制子蛋白。这些抑制子足以在每个细胞中从噬菌体基因组上刺激进一步的抑制子合成（图 17-27）。在多细胞生物体的发育过程中许多基因调控就是用这种方式工作的。我们将在第 19 章看到一些例子。

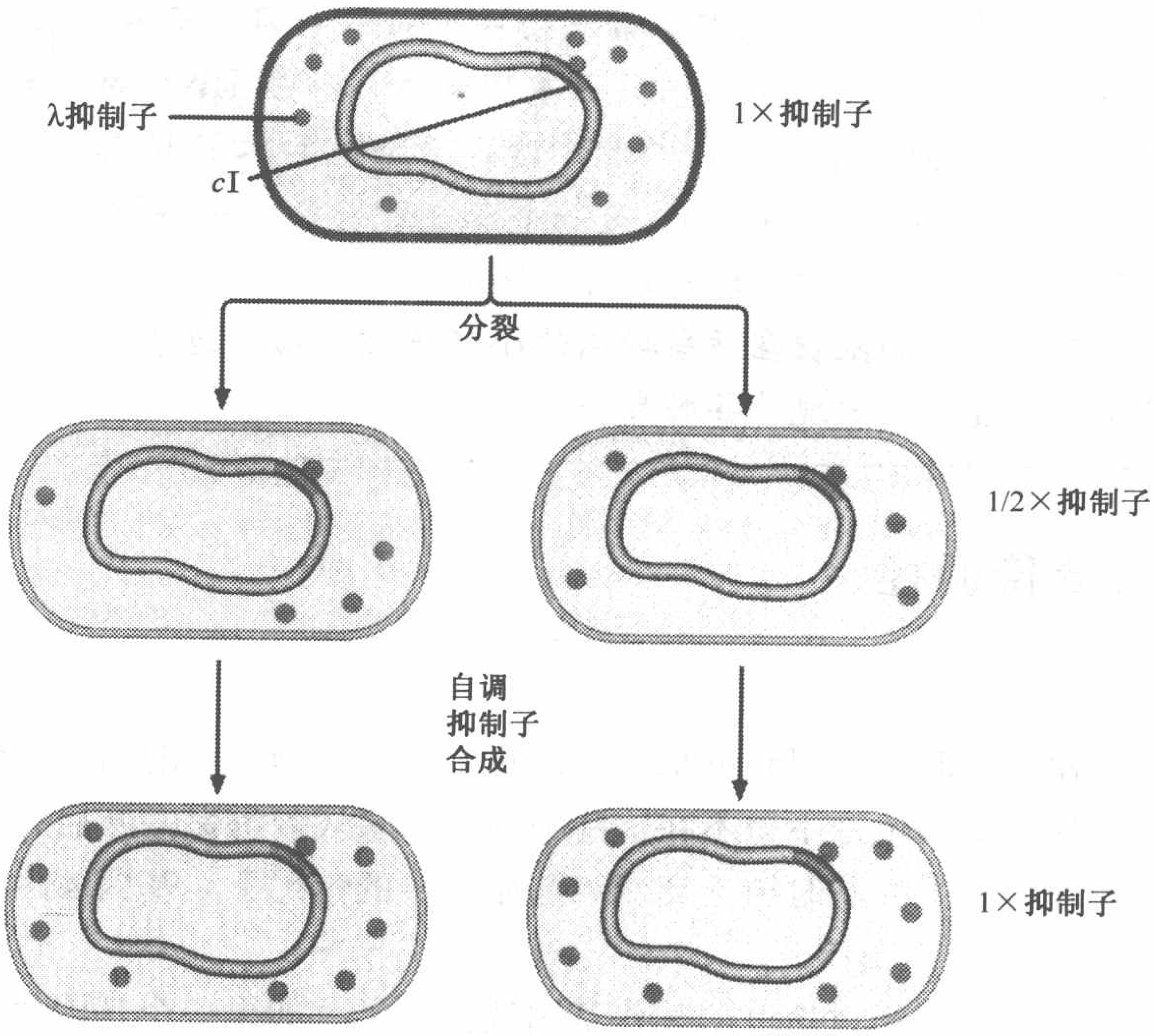


图 17-27　溶源性状态维持的表观遗传控制。

另一个著名的表观遗传调控的机制源于DNA甲基化，在对印记的描述中我们见过这样一个例子。DNA甲基化轻易跨越细胞分裂继承下来，如图17-28所示。如此，某些特定的DNA甲基化酶可以低频地甲基化先前未修饰的DNA；但是所谓的维持甲基化酶（maintenance methylases）极其有效地修饰半甲基化的DNA（即完全甲基化的DNA经复制后提供的那个底物）。在哺乳动物细胞中，DNA甲基化可能是沉默的基因组区域的主要标记。DNA复制后，在两个子细胞中的半甲基化位点被重新甲基化。它们随后可以被抑制子MeCP2所识别，然后后者募集组蛋白脱乙酰酶和甲基化酶，重新建立沉默（图17-25）。

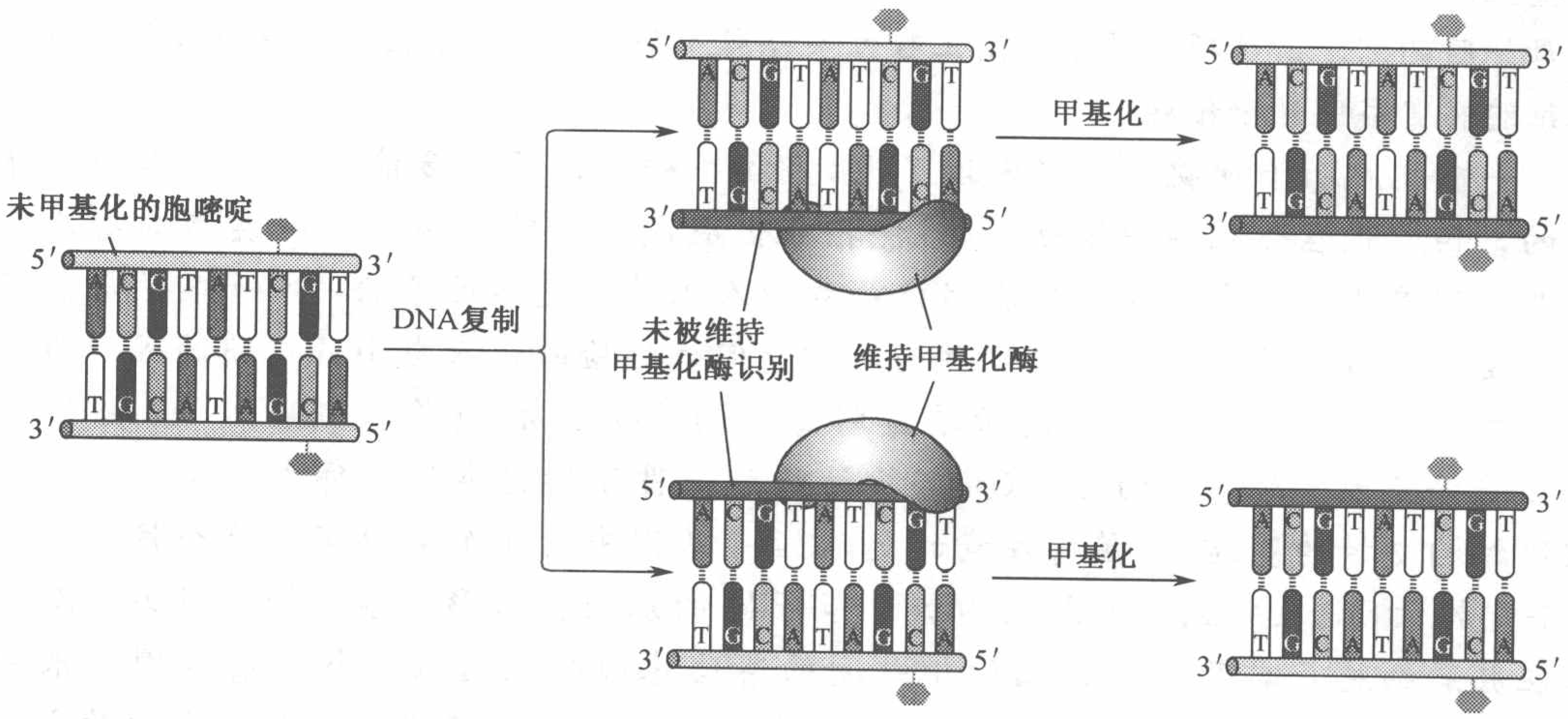


图17-28 DNA甲基化模式可以通过细胞分裂得以维持。如图17-25所示，参与脊椎动物基因表达的DNA可以被甲基化，进而该基因的表达被关闭。这种初始的甲基化由原位（de novo）甲基化酶执行。为了使某个处于关闭状态的基因永远关闭，甲基化状态必须通过细胞分裂得以遗传。此图显示了这是如何实现的。图中显示的DNA序列含有两个胞嘧啶，分别位于两条DNA链上，其中一个胞嘧啶已被甲基化，另一个则没有。这种模式在细胞分裂中得以维持，因为一旦DNA被复制，一种维持性的甲基化酶识别半甲基化的DNA，随后向其中的未被甲基化的胞嘧啶上加上甲基基团。完全未被甲基化的序列则不能被该酶识别，仍然保持非甲基化状态。所以，两个子DNA双螺旋链的甲基化模式最终与母链相同。（来源：Adapted，with permission from Alberts B. et al. 2002. Molecular biology of the cell，4th edition，p. 431，Fig. 7-81. © Garland science/Taylor & Francis LLC.）

核小体修饰在理论上可以为表观遗传的继承提供基础，尽管还没有发现这样的例子。考察被一段甲基化的组蛋白关闭的一个基因。当染色体的那个区域在细胞分裂过程中被复制，来自父本DNA分子的甲基化组蛋白最后在两个子代双链体间均匀分布（图7-42），于是每个子代分子携带一些甲基化的和一些非甲基化的核小体。甲基化的核小体可以募集带有染色质结构域的蛋白，包括组蛋白甲基化酶本身，它随后可以甲基化邻近的非修饰的核小体。这样，使用达到扩散的同样策略，染色质修饰的状态得以跨越几代而维持下来。尽管这是一个吸引人的模型，还要看看在缺乏DNA甲基化、DNA结合蛋白、或调控RNA时是如何工作的。

在框17-6展示的转录因子在决定基因表达模式的威力的最后一个例子中，我们将描述最近的一些实验，4个转录因子将一个体细胞转变成多能态。

框 17-6 用转录因子将体细胞重新编程为胚胎干细胞

用组织移植治疗某些人类疾病（如帕金森病、脊髓损伤、糖尿病）的一个主要障碍是病人免疫系统引发的对移植物的排斥。绕过这个排斥问题的一种方法是使用病人自己的体细胞产生胚胎干细胞。这些胚胎干细胞具有分化成体细胞的能力，因而为移植治疗提供一个量身定做的成体细胞来源（框 19-6）。直到最近，将一个体细胞转化为胚胎干细胞的唯一方法是体细胞核转移，这是一种将体细胞的核容物引入去核了的卵母细胞中的一个过程。然而，体细胞核转移在人类实验中被证实存在技术困难，而且已经引发了道德上的反对。最近，开发了一个令人兴奋的新技术，重新编程体细胞，它承诺绕过对卵母细胞的需求。这项新技术利用工程方法合成 4 个关键的转录因子来逆转分化，并将体细胞转变成胚胎干细胞。

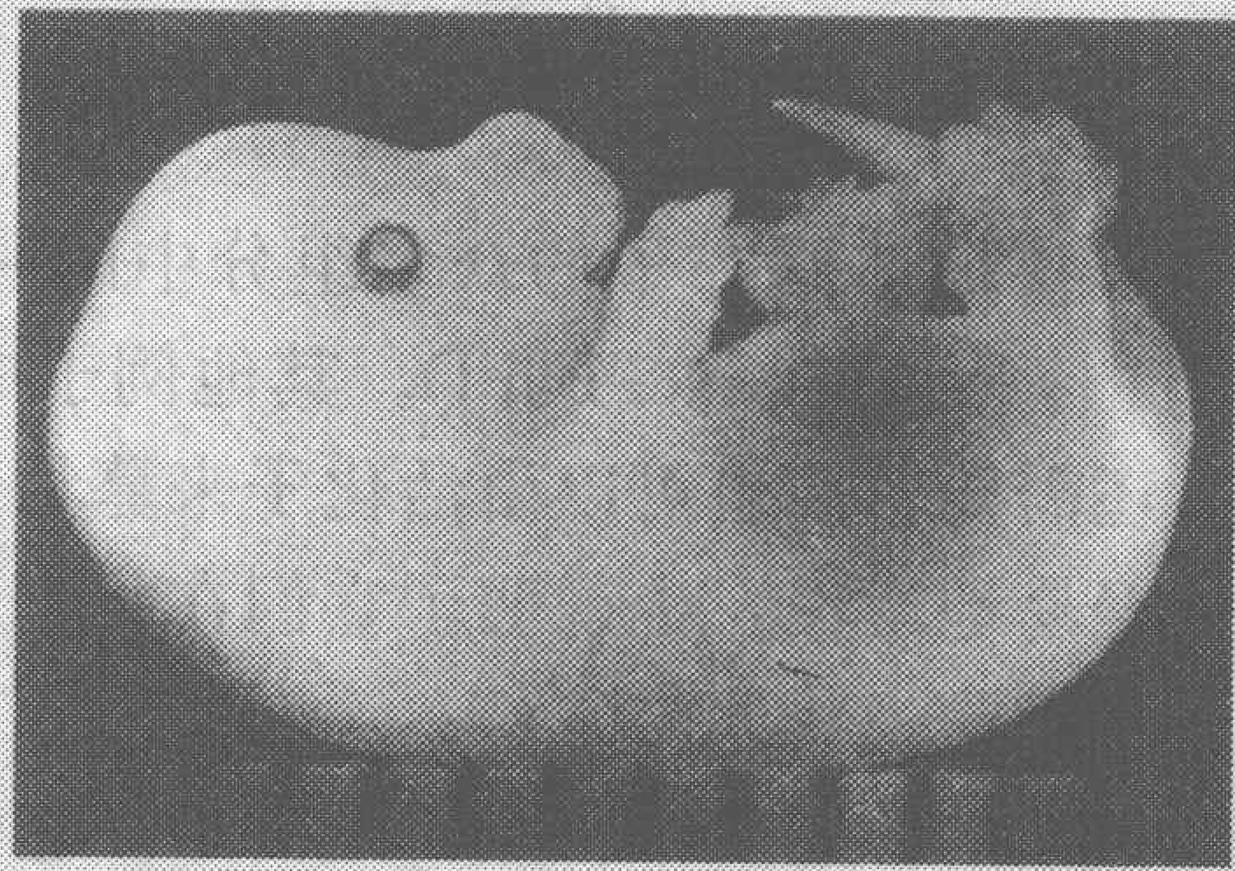
先前在胚胎干细胞的研究发现了大约 24 个转录因子是多能状态所必需的或相关的，但它们在体细胞中缺乏。（多能指分化成成体动物中许多细胞类型的潜能。）这些发现产生了一个假说，即在成体细胞中人工造成这些因子的一个子集的合成可以激发重新编程为一个多能态。作为对这个假说的检验，来自小鼠皮肤的成纤维细胞被工程改造（使用一个逆转录病毒将基因的 CDNA 拷贝导入细胞），来表达 24 个候选转录因子的各种组合。（成纤维细胞是一种体细胞类型，负责产生一种提供组织结构支持的基质。）值得注意的是，24 个基因中的 4 个足以实现这个目的。这 4 个是转录因子 Oct4、Sox2、c-Myc、和 Klf4 的基因。从形态学、增殖能力、及已知在胚胎细胞开启而在分化细胞中关闭的某些基因的表达（如 Nanog 基因）来判断，被工程改造后表达这 4 个基因的成纤维细胞看上去已经进入了胚胎类似的干细胞状态。

这些表达 Oct4、Sox2、c-Myc、和 Klf4 的细胞展现胚胎干细胞的其他特征吗？多能细胞的一个惊人的特征是，处于参与发育的关键调控基因的非编码控制元件的染色质展现一种已知为二价结构域的独特的修饰模式。二价结构域以在组蛋白 H3 的 Lys-4 上一大段甲基化和在 lys-27 上较小段的甲基化（第 7 章，尤其表 7-7）。已知 Lys-4 甲基化通过募集重塑酶和组蛋白乙酰化酶来启动转录，而在组蛋白 H3 的 Lys-27 上的甲基化通过染色质凝聚来抑制转录。（显然，这些二价结构域把持发育调控基因沉默，而同时又使它们做好被开启的准备。）当检测被 Oct4、Sox2、c-Myc、和 Klf4 重新编程的成纤维细胞，在纯正的成纤维细胞中主要展现要么 Lys-4 甲基化要么 Lys-27 甲基化的控制元件的染色质上，被发现展现了具有胚胎干细胞特征的二价模式。

尽管有这项惊人的结果，真正的多能需要展示重新编程细胞能够发育成一个成体动物。这个指标性的验证还没有满足，但是已经可能制造嵌合小鼠，其重新编程的细胞对所有成体细胞类型都做出了贡献，并且也已经可能生成完成来源于工程改造的成纤维细胞的小鼠晚期胚胎（框 17-6 图 1）。

表达 4 个特定基因足以重新编程一个成体细胞成为一个多能细胞的发现无疑是干细胞研究中的一个突破。但是，使用一个逆转录病毒和工程表达一个已知的癌症基因(Myc）排除了这项技术在人类治疗中的应用。然而，在将来，应该可能揭示从 Oct4、

Sox2、c-Myc 和 Klf4 到多能的一连串事件。如果这样的话，那么基于这些知识，有可能发现以一种不依赖工程表达潜在有害基因的良性方式来激发重新编程的药物。



框 17-6 图 1　全部源自于表达胚胎干细胞转录因子 Oct4、Sox2、c-Myc 和 Klf4 的成纤维细胞的晚期小鼠胚胎。（来源：Reprinted，with permission，from wernig M. et al. 2004. Nature 448：318-324，Fig. 5e. ©Macmillan.）

小结

如果不考虑生物体中那些可以被调控的附加步骤，那么就如同在细菌中一样，转录起始步骤是真核生物基因表达中最频繁被调控的步骤。也和在细菌中一样，转录起始步骤典型的是通过蛋白质来调控的，这些蛋白质结合在特定的靠近基因的 DNA 序列上，并能够使基因或者开启（活化子）或者关闭（抑制子）。尽管真核生物基因，存在原核生物中没有的在构造和转录上的复杂度（下面会总结），但这些调控机制的保守性还是被保留下来了。

核小体及其修饰。真核细胞中的 DNA 被组蛋白缠绕而形成了核小体。因此，转录机器和调控蛋白所结合的 DNA 序列在许多时候都被遮蔽了。修饰组蛋白的酶，通过增加（或移除）小的化学基团，在以下两个方面改变了组蛋白：它们改变了核小体被包裹的紧密度（因此也改变了核小体中的 DNA 的可接近性）；另外它们形成（或消除）了其他参与基因转录的蛋白质的结合位点。其他酶“重塑”了核小体：它们利用从 ATP 水解中得到的能量来移动周围的核小体，影响哪些序列被暴露出来。

许多的调控子和更远的调控距离。与对应的细菌相比较，多细胞真核生物的基因受控于更多的调控蛋白，一些蛋白质结合的位置与基因相距甚远。这反映出在多细胞生物中更多数量的生理信号参与了控制一个典型基因。

复杂精致的转录机制。RNA 聚合酶在细菌和真核生物中非常保守（第 12 章）。但是真核生物的酶类含有更多的亚基，而且在典型的真核生物启动子上除了聚合酶还结合有大约 50 个额外的蛋白质。这些蛋白质中有很多是以大的蛋白质复合体形式结合到启动子上。

在真核生物中，就如我们在细菌中所看到的，活化子主要通过募集来工作。在这些生物体中，活化子并不直接地或单独地募集聚合酶。它们募集其他启动某一特定基因的转录需要的蛋白复合体。RNA 聚合酶本身就是伴随着其他的这类复合体而被带入的。活化子也能够募集组蛋白修饰酶，而这些修饰产生的影响可以帮助转录机器结合到启动子上或起始有效的转录。

活化子能够和转录机器或核糖体修饰子的众多组分中的一个或更多个组分相互作用。例如，Gal4 募集中介蛋白，SAGA 和 TFⅡD。其他例子中，有效起始或延伸所需的因子可能在聚合酶结合后还需要它们也可以由活化子来募集。这解释了它们如何在大数量、多种结合方式下如此轻易地运作，也说明了它们在我们所知的信号整合及组合控制上的广泛运用，特别是在多细胞生物体中。

一些活化子从距离基因很远的位置发挥作用，这需要在它们的结合位点和启动子间的 DNA 形成环状突起。环是如何跨越在一些情况中非常远的距离，而形成仍不清晰，但是可能涉及活化子结合位点和启动子间的染色质的结构改变，使得那两个元件靠近。被称为绝缘子的 DNA 序列结合在蛋白质上，该蛋白质干扰限制了远距离增强子上的活化子与其启动子间的相互作用。这些绝缘子可以通过抑制那些使环化更容易的机制（如改变染色质结构）来发挥作用。绝缘子确保了活化子只对正确的基因产生效应。

真核抑制子通过各种途径工作，就如它们在细菌中一样。然而，细菌中可见的最简单和最普通的机制是抑制子结合到与启动子重叠的位点上，从而阻碍 RNA 聚合酶的结合。此机制并不常见于真核生物。最普通的是，真核抑制子通过募集可降低转录水平的组蛋白修饰物的途径来工作。举个例子，组蛋白乙酰基转移酶通常都与活化关联，而组蛋白脱乙酰酶——作为一种能够去除乙酰基团的酶，在这里产生抑制基因的效应。

在某些时候，核小体中一大段 DNA 部分能够通过适当的核小体修饰而被保留在相对惰性的状态，最显著的修饰就是去乙酰作用和甲基化。用这种方法，一组基因将能够保持在“沉默”状态而不需要在每个独立的基因上都有专有的抑制子。该状态一旦被建立就可以维持下去，因为修饰酶自身经常优先被募集到在该状态下的核小体中。从而，修饰状态募集了酶，而该酶产生了那个特定的修饰状态。这意味着一旦启动，沉默状态能够相当容易地被延续和传承。

在某些真核生物体中，如哺乳动物，沉默基因也和甲基化的 DNA 相关联。甲基化的序列可以阻碍转录机器与活化子的结合，或特异地结合一类抑制子，后者募集组蛋白修饰酶来抑制邻近的基因。DNA 甲基化在细胞分裂过程中可以保持（在既没有突变也没有起始该模式的信号存在的情况下），也就保持了被那些甲基化控制的基因表达模式。

如果一个基因的表达在细胞分裂过程中的某些阶段被保持，就被称为表观遗传。有各种机制使之实现。λ 溶源体的继承就是一个例子（依靠从启动子起始表达一个抑制子，随后表达变成自持的）。DNA 甲基化会影响基因表达，且容易继承，如我们已经看到的。通过使用组蛋白修饰的表观遗产通常作为一种可能性被讨论，但它们在没有 DNA 结合蛋白、调控蛋白、或 DNA 甲基化参与的情况下发挥这个作用的程度仍然不清楚。

（张清润　译　刘国振　吴琳　刘斌　校）

参考文献

书

Allis C.D., Jenuwein T., Reinberg D., and Caparros M.-L., eds. 2007. *Epigenetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Carey M. and Smale S.T. 2000. *Transcriptional regulation in eukaryotes: Concepts, strategies, and techniques*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Ptashne M. and Gann A. 2002. *Genes and signals*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

DNA识别

Garvie C.W. and Wolberger C. 2001. Recognition of specific DNA sequences. *Mol. Cell.* **8:** 937–946.

Harrison S.C. 1991. A structural taxonomy of DNA-binding domains. *Nature* **353:** 715–719.

激活

Bjorklund S. and Gustafsson C.M. 2005. The yeast Mediator complex and its regulation. *Trends Biochem. Sci.* **30:** 240–244.

Bulger M. and Groudine M. 2002. TRAPping enhancer function. *Nat. Genet.* **32:** 555–556.

Fry C.J. and Peterson C.L. 2001. Chromatin remodeling enzymes: Who's on first? *Curr. Biol.* **11:** R185–R197.

Jones K.A. and Kadonaga J.T. 2000. Exploring the transcription-chromatin interface. *Genes Dev.* **14:** 1992–1996.

Kim Y.J. and Lis J.T. 2005. Interactions between subunits of *Drosophila* Mediator and activator proteins. *Trends Biochem. Sci.* **30:** 245–249.

Kornberg R.D. 2005. Mediator and the mechanism of transcriptional activation. *Trends Biochem. Sci.* **30:** 235–239.

Lefstin J.A. and Yamamoto K.R. 1998. Allosteric effects of DNA on transcriptional regulators. *Nature* **392:** 885–888.

Malik S. and Roeder R.G. 2005. Dynamic regulation of pol II transcription by the mammalian Mediator complex. *Trends Biochem. Sci.* **30:** 256–263.

Myers L.C. and Kornberg R.D. 2000. Mediator of transcriptional regulation. *Annu. Rev. Biochem.* **69:** 729–749.

Naar A.M., Lemon B.D., and Tjian R. 2001. Transcriptional coactivator complexes. *Annu. Rev. Biochem.* **70:** 475–501.

Ptashne M. and Gann A. 1997. Transcriptional activation by recruitment. *Nature* **386:** 569–577.

Struhl K. 1999. Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell* **98:** 1–4.

抑制

Maldonado E., Hampsey M., and Reinberg D. 1999. Repression: Targeting the heart of the matter. *Cell* **99:** 455–458.

Smith R.L. and Johnson A.D. 2000. Turning genes off by Ssn6-Tup1: A conserved system of transcriptional repression in eukaryotes. *Trends Biochem. Sci.* **25:** 325–330.

核小体修饰因子及转录调控

Berger S.L. 2002. Histone modifications in transcriptional regulation. *Curr. Opin. Genet. Dev.* **12:** 142–148.

Flaus A. and Owen-Hughes T. 2001. Mechanisms for ATP-dependent chromatin remodeling. *Curr. Opin. Genet. Dev.* **11:** 148–154.

Jenuwein T and Allis C.D. 2001. Translating the histone code. *Science* **293:** 1074–1080.

Marmorstein R. and Roth S.Y. 2001. Histone acetyltransferases: Function, structure, and catalysis. *Curr. Opin. Genet. Dev.* **11:** 155–161.

Narlikar G.J., Fan H.Y., and Kingston R.E. 2002. Cooperation between complexes that regulate chromatin structure and transcription. *Cell* **108:** 475–487.

Peterson C.L. and Workman J.L. 2000. Promoter targeting and chromatin remodeling by the SWI/SNF complex. *Curr. Opin. Genet. Dev.* **10:** 187–192.

沉默、印迹及表观遗传学

Bird A.P. and Wolffe A.P. 1999. Methylation-induced repression—Belts, braces, and chromatin. *Cell* **99:** 451–454.

Gartenberg M.R. 2000. The Sir proteins of *Saccharomyces cerevisiae:* Mediators of transcriptional silencing and much more. *Curr. Opin. Microbiol.* **3:** 132–137.

Goldberg A.D., Allis C.D., and Bernstein E. 2007. Epigenetics: A landscape takes shape. *Cell* **128:** 635–638.

Gottschling D.E. 2000. Gene silencing: Two faces of SIR2. *Curr. Biol.* **10:** R708–R711.

———. 2004 Summary: Epigenetics—From phenomenon to field. *Cold Spring Harbor Symp. Quant. Biol.* **69:** 507–519.

Grunstein M. 1998. Yeast heterochromatin: Regulation of its assembly and inheritance by histones. *Cell* **93:** 325–328.

Klose R.J. and Bird A.P. 2005. Genomic DNA methylation: The mark and its mediators. *Trends Biochem. Sci.* **31:** 89–97.

Ptashne M. 2007. On the use of the word "epigenetic." *Curr. Biol.* **17:** R233–R236.

Richards E.J. and Elgin S.C. 2002. Epigenetic codes for heterochromatin formation and silencing: Rounding up the usual suspects. *Cell* **108:** 489–500.

Tilghman S.M. 1999. The sins of the fathers and mothers: Genomic imprinting in mammalian development. *Cell* **96:** 185–193.

Wolffe A.P. 2000. Transcriptional control: Imprinting insulation. *Curr. Biol.* **10:** R463–R465.

Wood A.J. and Oakey R.J. 2006. Genomic imprinting in mammals: Emerging themes and established theories.

PloS Genet. **2:** 1677–1685.

组合控制与协同

Arnosti D.N. and Kulkarin M.M. 2005. Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *J. Cell Biol.* **94:** 890–898.

Carey M. 1998. The enhanceosome and transcriptional synergy. *Cell* **92:** 5–8.

Johnson A.D. 1995. Molecular mechanisms of cell type determination in budding yeast. *Curr. Opin. Genet. Dev.* **5:** 552–558.

Maniatis T., Falvo J.V., Kim T.H., Kim T.K., Lin C.H., Parekh B.S., and Wathelet M.G. 1998. Structure and function of the interferon-β enhanceosome. *Cold Spring Harb. Symp. Quant. Biol.* **63:** 609–620.

Merika M. and Thanos D. 2001. Enhanceosomes. *Curr. Opin. Genet. Dev.* **11:** 205–208.

Rokas A. 2006. Evolution: Different paths to the same end. *Nature* **443:** 401–402.

大范围互作

Dean A. 2006. On a chromosome far, far away: LCRs and gene expression. *Trends Genet.* **22:** 38–45.

Dorsett D. 1999. Distant liaisons: Long-range enhancer–promoter interactions in *Drosophila*. *Curr. Opin. Genet. Dev.* **9:** 505–514.

Gaszner M.and Felsenfeld G. 2006. Insulators: Exploiting transcriptional and epigenetic mechanisms. *Nat. Rev. Genet.* **7:** 703–713.

Krueger C. and Osborne C.S. 2006. Raising the curtains on interchromosomal interactions. *Trends Genet.* **22:** 637–639.

Li Q., Barkess G., and Qian H. 2006. Chromatin looping and the probability of transcription. *Trends Genet.* **22:** 197–202.

信号及信号转导

Bromberg J.F. 2001. Activation of STAT proteins and growth control. *Bioessays* **23:** 161–169.

Brown M.S., Ye J., Rawson R.B., and Goldstein J.L. 2000. Regulated intramembrane proteolysis: A control mechanism conserved from bacteria to humans. *Cell* **100:** 391–398.

Darnell J.E., Jr. 1997. STATs and gene regulation. *Science* **277:** 1633–1635.

Hill C.S. and Treisman R. 1995. Transcriptional regulation by extracellular signals: Mechanism and specificity. *Cell* **80:** 199–211.

Hunter T. 2000. Signaling—2000 and beyond. *Cell* **100:** 113–127.

Pawson T. and Nash P. 2000. Protein–protein interactions define specificity in signal transduction. *Genes Dev.* **14:** 1027–1047.

抑制与疾病

Gabellini D., Green M.R., and Tupler R. 2004. When enough is enough: Genetic diseases associated with transcriptional derepression. *Curr. Opin. Genet. Dev.* **14:** 301–307.

Gabellini D., Tupler R., and Green M.R. 2003. Transcriptional derepression as a cause of genetic diseases. *Curr. Opin. Genet. Dev.* **13:** 239–244.

Kriaucianis S. and Bird A. 2003. DNA methylation and Rett syndrome. *Hum. Mol. Genet.* **12:** R221–R227.

Miller G. 2007. Medicine. Rett symptoms reversed in mice. *Science* **315:** 749.

第18章 调控 RNA

本章概要

- **细菌中 RNA 介导的调控**
- **RNA 干扰是真核生物的主要调控机制**
- **miRNA 分子的合成与功能**
- **RNAi 的进化和应用**
- **调控 RNA 与 X 染色体灭活**

前两章中，我们讨论了原核生物和真核生物的转录调控。我们知道这一调控是通过调控蛋白，尤其是序列特异性 DNA 结合蛋白来进行的。这些蛋白质能够激活或抑制其附近基因的转录。大约 50 年前，Francois Jacob 和 Jacques Monod 提出了他们的抑制子模型（框 16-2），由此开始了基因调控的详细机制的研究。而那时，他们不能告诉我们反式因子（*trans* factor），即抑制子（repressor），究竟是蛋白质还是 RNA。我们得知，他们所研究的，实际上还有很多其他研究者所研究的这些调控因子都是蛋白质。这些蛋白质通过结合到 DNA 的操纵子（operator）位置上而起作用。但是，在他们的原文中，他们轻易地相信这些调控因子可能是 RNA 分子——他们的确更倾向于这种可能性。

随着越来越多的蛋白质调控因子在原核和真核生物中被发现，RNA 分子可能是调控因子这一观点几乎被遗忘。但在近几年，RNA 调控子的研究获得了新的突破。尤其在真核生物中，RNA 分子被发现能够在转录水平尤其是在翻译水平上发挥调控作用。这一新研究领域源于如下两点发现：最先报道于 20 世纪 90 年代初的小 RNA（microRNA）的发现，和 90 年代末 RNA 干扰（RNA interference）现象的发现。在描述这种新的调控形式之前，包括它们是如何作用以及能给研究者带来何种应用价值，我们先来考虑最先在细菌中发现的 RNA 介导的基因调控。

细菌中 RNA 介导的调控

小 RNA 存在于原核生物之中在多年前就已成为一个不争的事实。这些 RNA，其中一部分参与质粒的复制调控过程，另一部分则与基因表达调控相关（参看第 11 章中有关 Tn*10* 的讨论）。对于后者，它们有些能够参与转录调控过程，如 *E. coli* 的 6S RNA。通过结合到 RNA 聚合酶的 σ^{70} 亚基上，6S RNA 能够下调大多数

σ^{70}启动子控制的转录。细菌处于静止期时（营养物质耗尽时，细菌进入一个特殊的生长阶段，此时细胞停止分裂，详见第 22 章），6S RNA 保持很高的浓度。静止期还产生另一种不同的 σ 因子，σ^{S}。σ^{S} 因子与 σ^{70} 竞争与核心聚合酶的结合，并引导聚合酶到达特定的启动子，促进许多压力应答性基因的表达，使细菌能在静止期存活。通过下调 σ^{70} 启动子的转录，6S RNA 有助于转而表达 σ^{70} 启动子调控的基因的表达。

近年来，研究的焦点转移到另一组细菌的小 RNA 分子上。这些小 RNA 分子能够调节 mRNA 翻译和降解。由于它们与真核生物中调控基因表达的 RNA（包括小分子干扰 RNA 和 microRNA，我们将会在本章第二部分讨论）相似，人们对这种小 RNA 研究的兴趣进一步高涨。与真核生物的调控 RNA（长 21～30 核苷酸）相比，细菌小 RNA 要大得多（称为 sRNA，长 80～110 核苷酸），并且是由小基因直接编码形成其最终产物，而不是像真核生物 RNA 调控子那样，是通过大分子双链 RNA 的前体加工而形成的。很多这些基因已通过生物信息学被鉴定。在 *E. coli* 中约有 100 个，其中十来个 sRNA 的特性已经被阐述。大多数 sRNA 与靶 mRNA 的互补序列碱基配对而引导其降解从而抑制翻译，但在一些特殊例子中反而能刺激翻译。

在大多数情况下，细菌 Hfq 蛋白辅助 sRNA 结合到靶 mRNA 上。这些 RNA 伴侣是不可或缺的。由于 sRNA 和靶 mRNA 之间的互补很不完全，且互补的序列很短，因此它们之间的相互作用较弱。Hfq 有助于碱基配对，并且在 sRNA 与靶 mRNA 配对之前，Hfq 先与 sRNA 结合，从而提高了这些调控因子的稳定性。

大肠杆菌中研究得较为深入的 sRNA 是长 81 个核苷酸的 sRNA（RybB RNA）。该 sRNA 能够与好几个靶 mRNA 结合而引发其降解，因为配对形成的异源双链核酸分子能被核酸酶 RNase E 识别并当成底物。与 RybB 作用的靶 mRNA，大多数是编码贮铁蛋白的。在特定环境中，游离铁离子是细胞所必需的。但是，高浓度的铁离子则是有毒的。RybB 通过调控贮铁蛋白的水平，在细胞中起到调节铁离子浓度的作用。调控 RybB 表达到启动子会被一种特异的 σ 因子（称为 σ^{E}）识别。同 σ^{S} 一样，σ^{E} 也是一种压力应答 σ 因子。编码 σ^{E} 的基因表达受 RybB 自身调控，所以这种 sRNA 是 σ^{E} 自身负调控环（autonegative regulatory loop）的组成部分。

如上所述，静止期 σ 因子 σ^{S} 由 *E. coli* 的 *rpoS* 基因所编码。*rpoS* mRNA 翻译过程受两种 sRNA 调节：DsrA 和 RprA。翻译激活是通过另外一种 RNA 碱基配对的开关机制来完成的：小 RNA 通过与 mRNA 的一个区域结合来抑制翻译，否则该区域将会与核糖体结合位点配对而促进翻译。*rpoS* 基因也受其他小 RNA 的负作用，如 OxyS。图 18-1 显示这两种机制。

我们将在本章的后半部分回过来讨论那些反式作用调控 RNA（*trans*-acting regulatory RNA）。在那里我们会对它们在真核生物基因表达调节方面的作用进行讨论。但在进入该主题之前，我们先剖析细菌中基因调控的另外一些例子。这种调控是由另外一种顺式（*cis*−）的 RNA 碱基配对介导的。这种顺式调控是指 RNA 调控元件调控的那些自身含有 mRNA 的基因的表达。此类调控的典型例子是那些调控代谢型操纵子（metabolic operon）和合成型操纵子衰减（attenuation in biosynthetic operon）的“核糖开

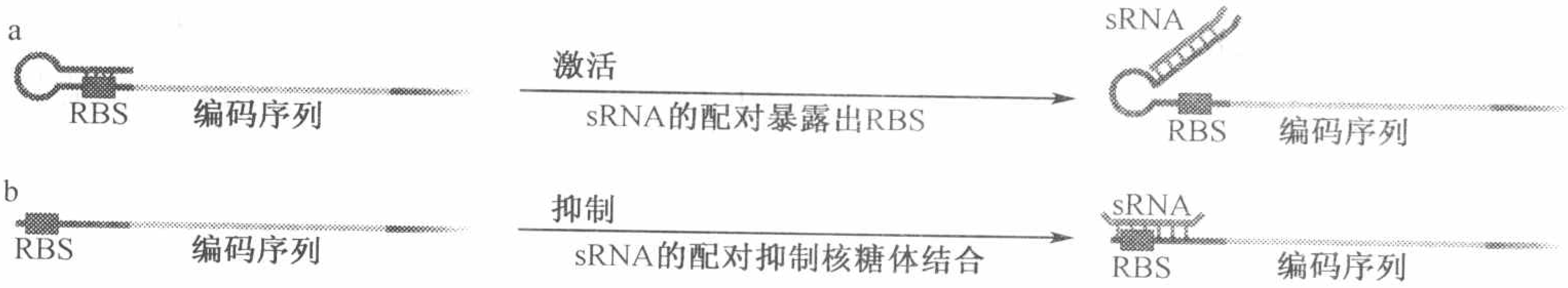


图 18-1 sRNA 对翻译的激活与抑制。 RBS 代表核糖体结合位置。如果它由于与另一个 RNA 分子（如 b 所示），或者与同一个 RNA 分子的另一个区域（如 a 所示）的碱基配对而被封闭，翻译就被抑制。（引自 Gottesman S. et al. 2006. Cold Spring Harbor Symp. Quant. Biol. 71：1-11，Fig. 1. © Cold Spring Harbor Laboratrory Press，经许可使用、修改）

关”（riboswitch）。*E. coli* 的 *trp* 基因便是后一机制的经典例子，RNA 介导的调控也是在该例子中首次被发现的（框 18-1 中将详细描述该例子）。

基因转录物中的核糖开关——通过改变二级结构调控基因表达

核糖开关是通过对小分子的浓度改变进行应答来调控基因的表达。这些调控元件特异地存在它们所调控的基因的 5′-非翻译区，可以在转录或翻译水平调控基因的表达。我们将会在下面讨论到，该调控是通过改变 RNA 的二级结构来完成的。

一个核糖开关都是由两部分组成的：适配体（aptamer）和表达平台（expression platform）（图 18-2）。适配体与小分子配基（ligand）结合，从而引起其构象改变，进而导致与其相邻的表达平台的二级结构也产生了变化。这种构象的变化通过终止基因转录或是抑制翻译的启动，改变了相关基因的表达。这两种机制在图 18-3 所示的例子中已有详细说明，我们现在对此进行描述。

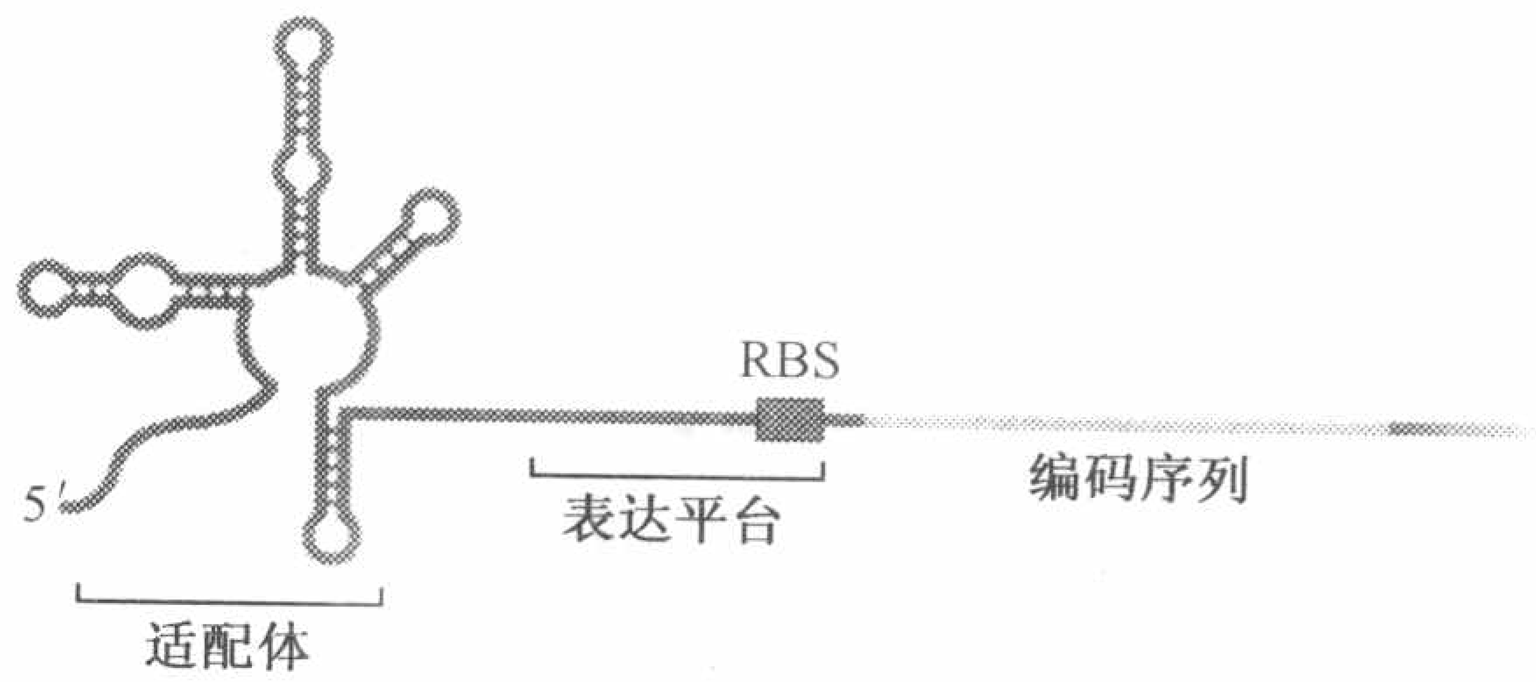


图 18-2 核糖开关 RNA 的结构。 如正文所描写的，适配体与控制代谢物结合，导致相连的表达平台的结构改变。迄今鉴定的适配体大小在 70 到 200 个核苷酸之间变动；而表达平台在大小和特征上都具有更大的不同。

毋庸置疑，核糖开关特异地存在于它所识别的代谢配基合成相关基因的上游。例如，在枯草杆菌（*Bacillus subtilis*）中，很多与甲硫氨酸代谢有关的基因都具有一个长 200 个核苷酸的非翻译前导 RNA 序列。这些 RNA 起到 SAM（腺苷甲硫氨酸）感受性核糖开关（S-adenosylmethionine-sensing riboswitch）的作用。RNA 聚合酶在启动子上启动转录，并从这段前导序列中开始，然后进入下游基因的编码序列。前导序列被转

录成 RNA 后，它通过内部碱基配对的不同的替换方式形成不同的结构（图 18-3a）。其中一种结构就是形成柄-环转录终止子（见第 12 章）。SAM（核糖开关的配体）与适配体结合，稳定包括转录终止子在内的二级结构（图 18-3a 下部所示）。在这种情况下，聚

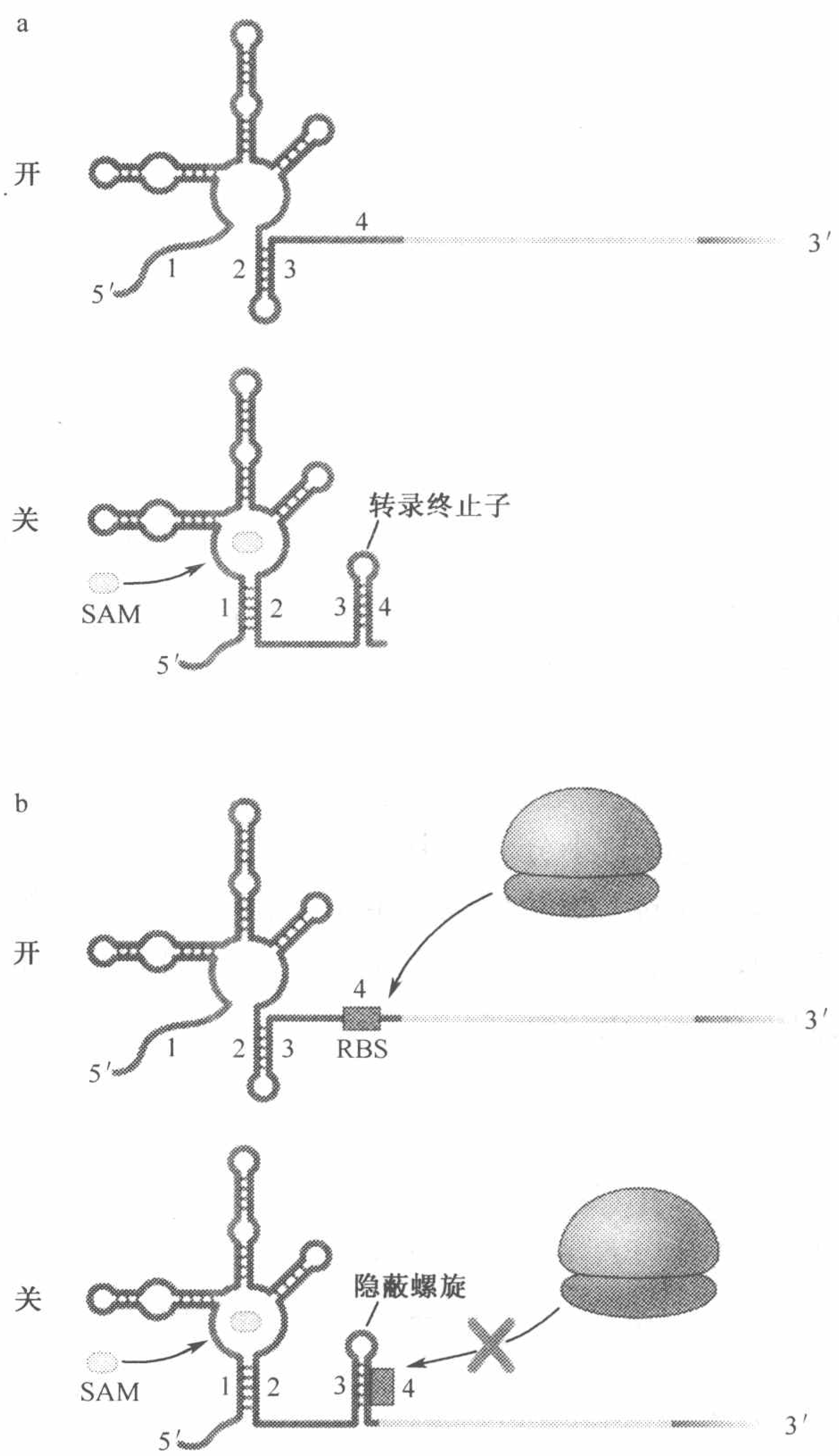


图 18-3 核糖开关调控转录终止或翻译起始。图示 SAM 感受性核糖开关的两个例子：a. 转录终止的调控，b. 翻译起始的调控。数字 1～4 表示 RNA 编码区上游不同的序列元件区域（以黄色表示）。没有 SAM 时，区域 2 和 3 形成柄-环；存在 SAM 时，区域 3 和 4 也同区域 1 和 2 一样都形成柄-环。如图所示，转录或者翻译的调控都是这种二级结构改变的结果。a. 区域 3 和 4 形成的柄-环产生了一个转录终止子，它使得 RNA 聚合酶在转录这一区域之后，进入下游编码区域之前立即终止转录。在这一例子中，mRNA 的柄-环后面是一串 U，这是转录终止子的另一个特征（图 12-12）。b. 图示，区域 3 和 4 形成的柄-环通过封闭核糖体结合位点而阻止了翻译的起始。

合酶在有机会转录下游基因的蛋白质编码序列之前，转录过程就被终止。这种转录调控的方式称为衰减作用（attenuation）。在另一基因即另一情况下，SAM 感受性核糖开关可以通过调控翻译过程起作用。如图 18-3b 所示，在该情况下，SAM 结合到适配子而稳定其替换二级结构，这一二级结构包含非转录终止子的柄-环结构，也包括一个核糖体结合位点区域（RBS，ribosme-binding site，图 18-3 中标以 4 的区域）。由 SAM 引起的这一构象改变，隔离了 RBS 并且阻止了核糖体使翻译不能进行。因此，这种方式的翻译抑制实质上与前文描述的反式作用 sRNA（*trans*-acting sRNA）作用机制相同（图 18-1）。图 18-4 显示了由 SAM 与核糖开关结合所引起的 RNA 二级结构改变的细节。

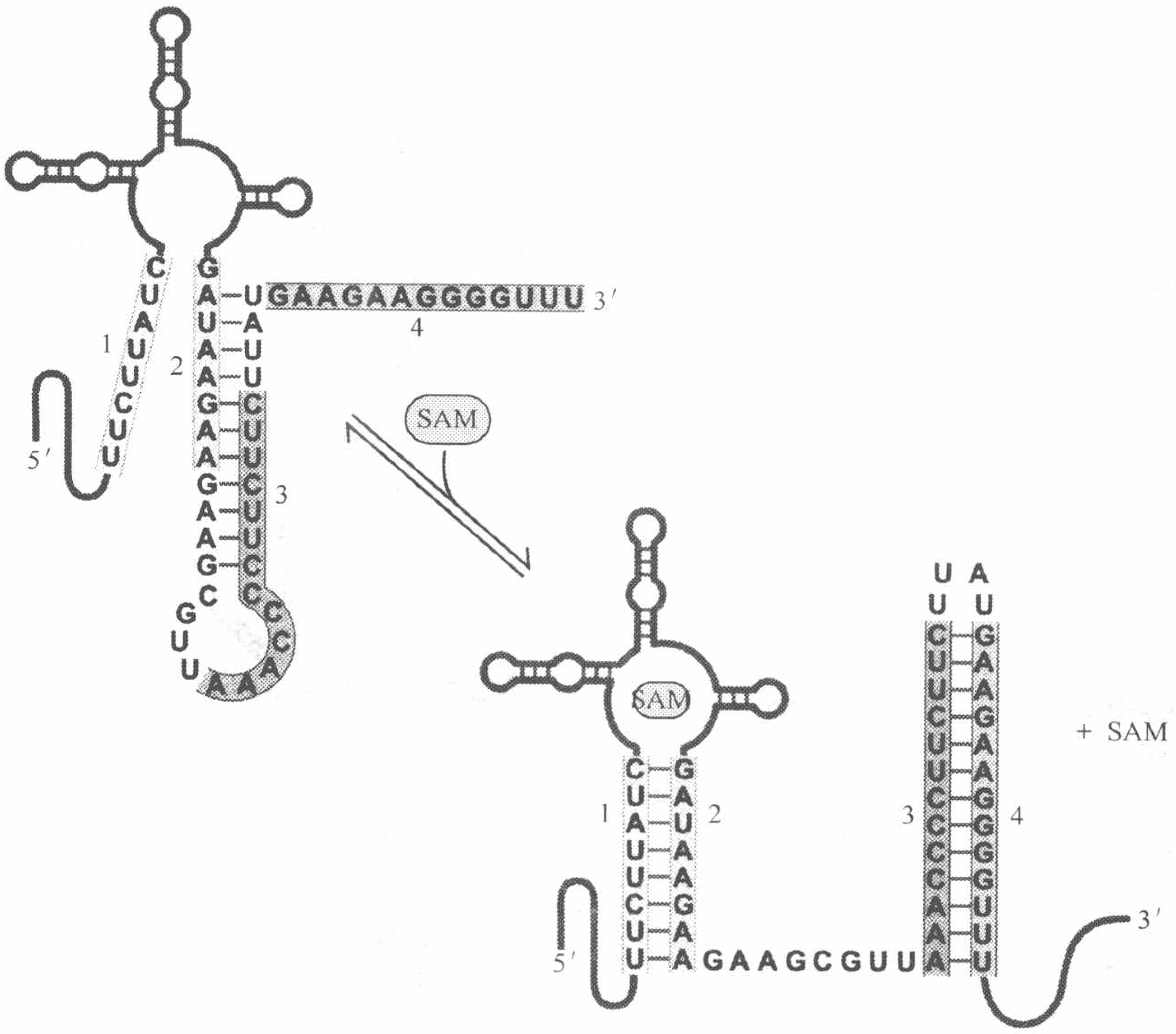


图 18-4 SAM 感受性核糖开关二级结构的改变。这里显示了区域 1～4（图 18-3）的详尽序列并用颜色标出。显示了有无 SAM 结合两种情况下，两个不同的二级结构的碱基配对。（据 Winkler W. C. et al. 2003. *Nat. Struct. Biol.* 10：701-707，Fig. 5b. ©Macmillan 修改。）

很多核糖开关已经被确认，而且它们能够对许多不同的代谢物发生应答。这些配体包括赖氨酸和其他氨基酸、维生素 B_{12}、辅酶焦磷酸硫胺素（coenzyme thiamine pyrophosphate，TPP），黄素单核苷酸（flavin mononucleotide，FMN）和鸟嘌呤（图 18-5）。

另一种核糖开关对空载 tRNA（uncharged tRNA）发生应答，而不是那些小分子配体。一些基因，尤其是氨酰-tRNA 合成酶基因（第 14 章），受长为 200～300 核苷酸的非翻译前导 RNA 通过衰减机制来调控。该前导 RNA 序列能够直接地、特异性地与合成酶的同类空载 tRNA 结合，而不能与负载 tRNA（charged tRNA）结合。负载 tR-

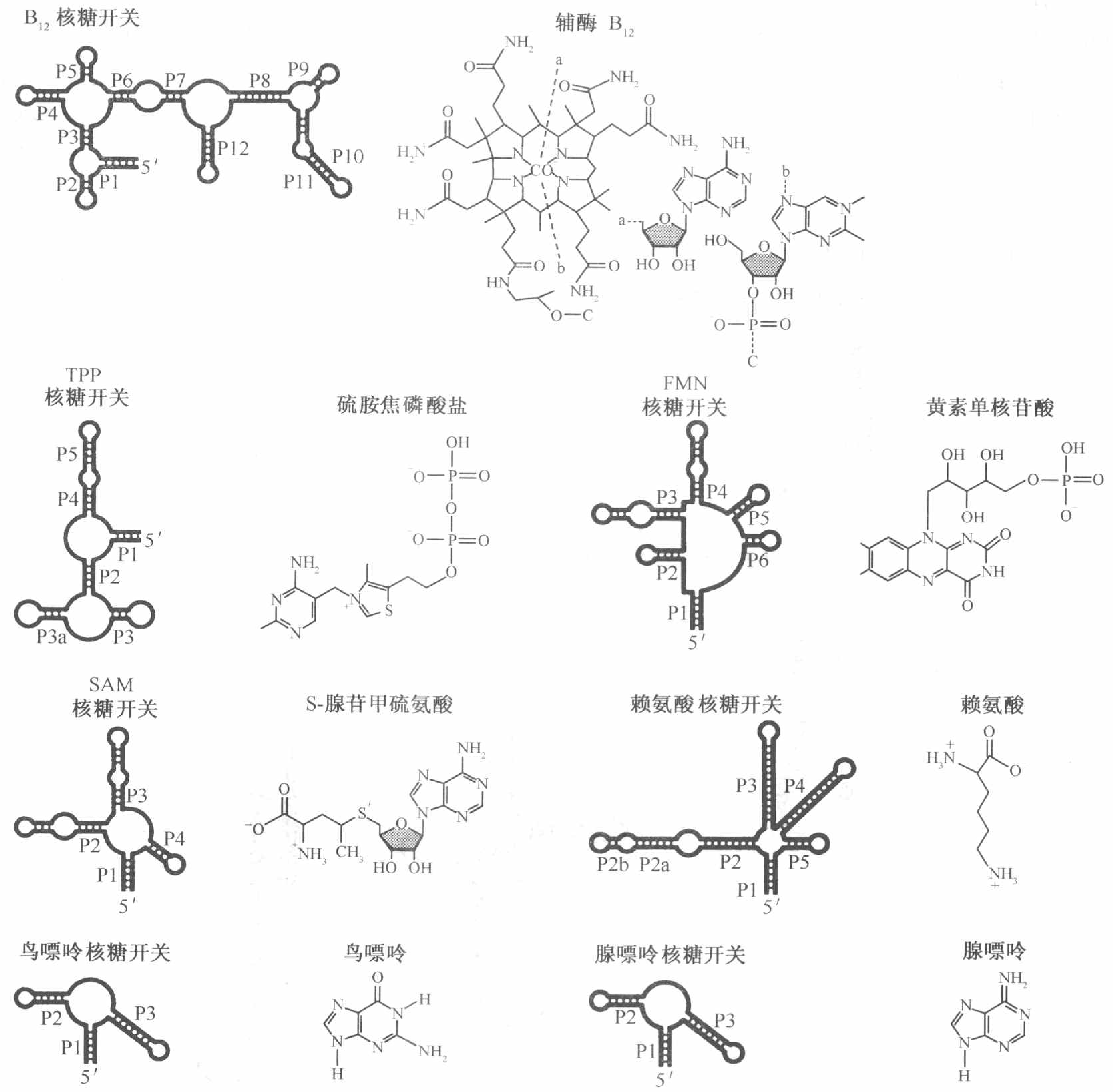


图 18-5 核糖开关对各种代谢物的应答。这里显示了 7 种核糖开关以及它们感受的代谢物。(经许可,引自 Mandal M. et al. 2003. *Cell* 113:557-586,Fig. 7A ©Elsevier.)

NA 不能同 RNA 二级结构所提供的结合域(binding pocket)相适应。通过与空载 tRNA 结合,前导 RNA 保持抗终止结构(antitermination structure),从而使得其临近的合成酶基因的转录得以继续。该调控的特异性是通过 tRNA 和前导 RNA 之间的"密码子-反密码子"相互作用形成的。由于空载 tRNA(不是负载 tRNA)可以与前导 RNA 结合,所以只有当细胞内同类氨基酸供应不足且空载 tRNA 的水平升高时,转录通读过程才会被激活。

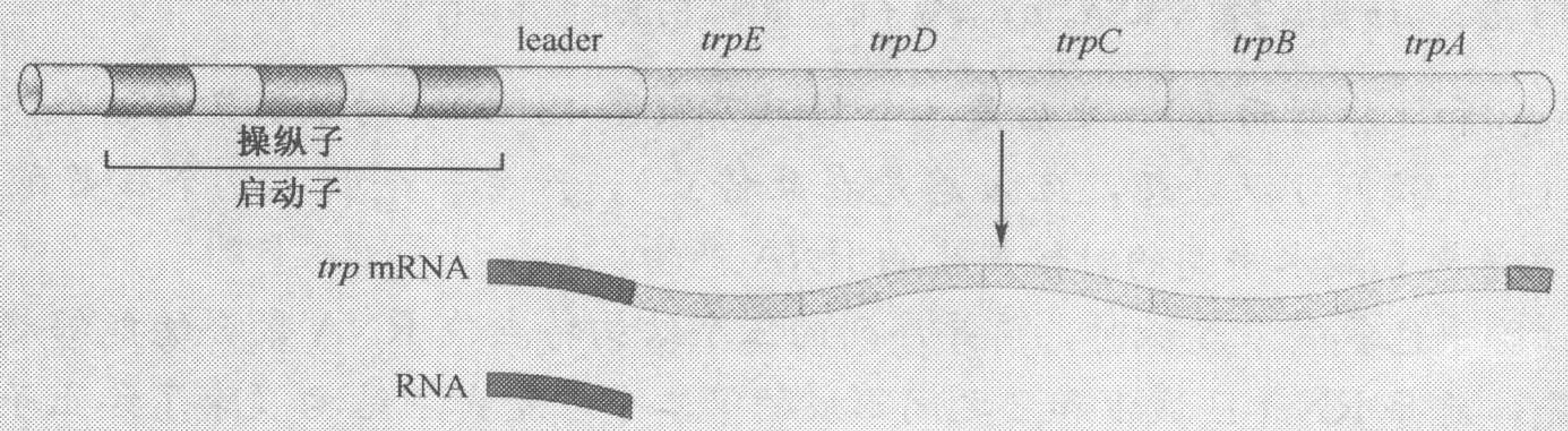
虽然在细菌中最为多见,核糖开关也已在其他生物,包括古菌、真菌和植物中被发现。在一些高等生物例子中,核糖开关甚至还参与替换剪接的调控(第 13 章)。譬如最近报道,在真菌粗糙脉孢霉(*Neurospora crassa*)中鉴定了三种 TPP 适配体,其中两

种适配体通过调节 RNA 剪接而抑制基因表达，而第三种却起激活作用。

两种 RNA 二级结构的替换所介导的衰减作用首先是在 *E. coli* 的色氨酸操纵子的研究中发现的（框 18-1）。色氨酸操纵子包括负责色氨酸生物合成的所有基因。这些基因的表达，是通过感受细胞中的待用色氨酸负载 tRNA（charged $tRNA^{trp}$）的水平，对细胞中色氨酸的浓度发生应答。这已成为众多相似情况的一个例子，和细胞中的待用氨基酸相关联的衰减作用，是氨基酸生物合成操纵子的一种普遍测控机制。色氨酸操纵子如何被调控的研究，首次证明了 RNA 二级结构的替换能够调控任何机体中基因的表达。

框 18-1　受衰减作用调控的氨基酸生物合成操纵子

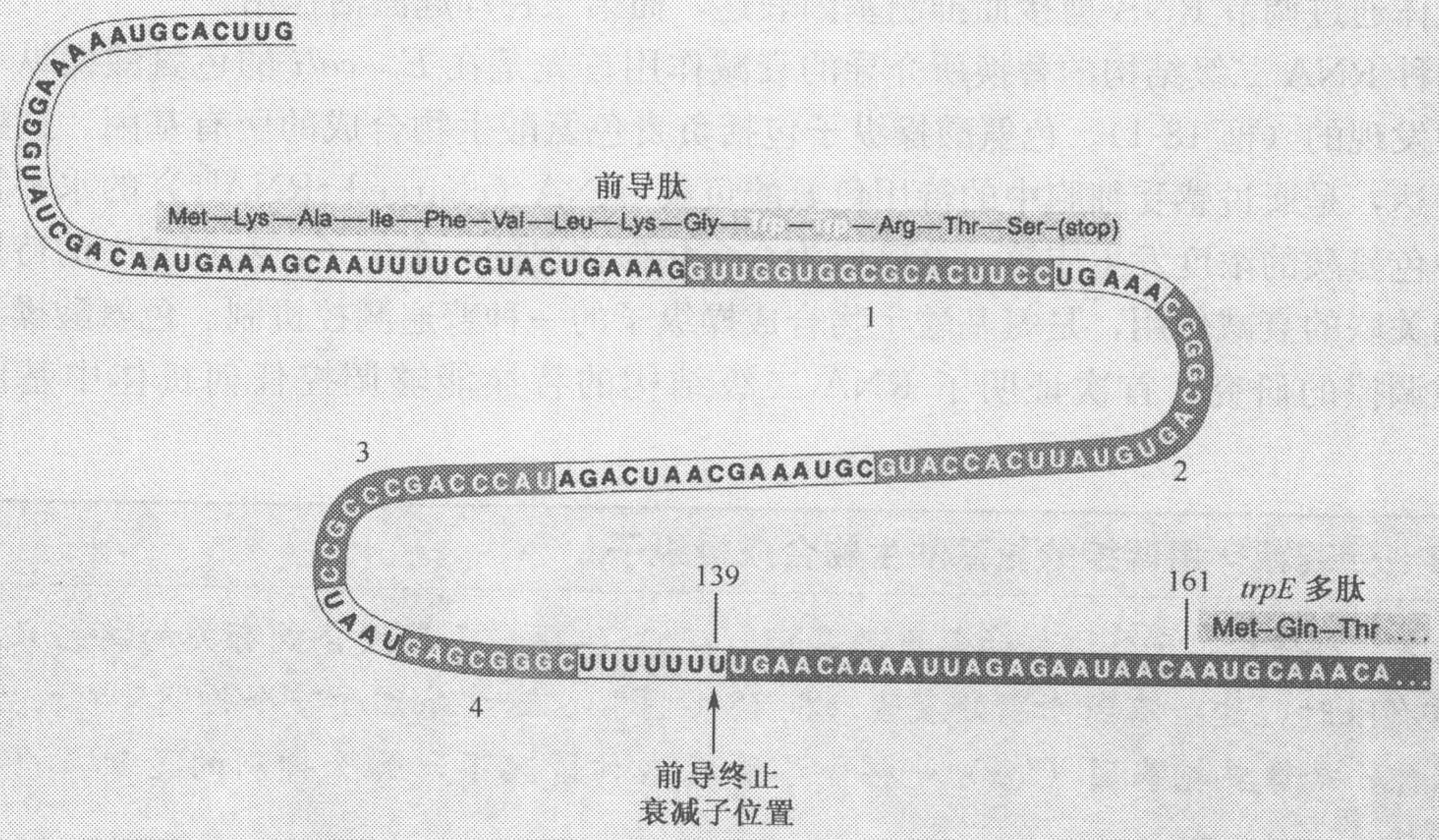
在 *E. coli* 中，5 个相邻的色氨酸基因（*trp*）编码与色氨酸合成有关的酶。只有当色氨酸有限时，这些基因才高效表达（框 18-1 图 1）。这些基因受一个抑制子（repressor）调控，就像乳糖基因（*lac*）一样。尽管在这种情况下，解除抑制的是配基（色氨酸）的缺乏。



框 18-1 图 1　*Trp* 操纵子。 *E. coli* 的色氨酸操纵子，图示前导 RNA（见正文）与编码 Trp 酶的结构基因之间的关系。基因产物为氨基苯甲酸盐合成酶（anthranilate synthetase，*trpE* 的产物），磷酸核糖基邻氨基苯甲酸转移酶（phosphoribosyl anthranilate transferase，*trpD* 的产物），磷酸核糖基邻氨基苯甲酸异构酶-吲哚甘油磷酸合成酶（phosphoribosyl anthranilate isomearase-indole glycerol phosphate synthetase，*trpC* 的产物），色氨酸合成酶 β（tryptophan synthetase，*trpB* 的产物），色氨酸合成酶 α（tryptophan synthetase，*trpA* 的产物）。

然而，即使 RNA 聚合酶启动了色氨酸 mRNA 分子的转录，它也并不总是能产生完整的转录物。就像核糖开关一样，产生完整的转录物的最后决定，要受衰减作用的控制。在这个例子中，大多数转录在未成熟前就终止，甚至在没转录完第一个色氨酸基因（*trpE*）之前就停止。但是，如果细胞中色氨酸水平很低的话，衰减作用会被克服；当色氨酸有限时，聚合酶不会终止转录，相反会转录出所有 5 个 *trp* 基因。就像核糖开关一样，衰减作用是否发生，依赖于 RNA 形成不同的替换二级结构的能力。然而在这个例子中，前导 RNA 选择形成两种不同的替换二级结构中的哪一种，并不受直接结合到前导 RNA 上的配基的控制。相反，这种选择依赖于细菌中转录与翻译之间的偶联（coupling）。

Trp 操纵子 mRNA 的 5′端序列含有一个长 161 核苷酸的前导序列，这一序列位于 *trpE* 第一个密码子的上游（框 18-1 图 2）。在这段前导序列之后，*trpE* 之前是一个转录终止子，RNA 中有一个特征性的环状的“发卡”（hairpin）结构（由框 18-1 图 2



框 18-1 图 2　*trp* 操纵子前导 RNA。*trp* 操纵子前导 RNA 的核苷酸序列特征。

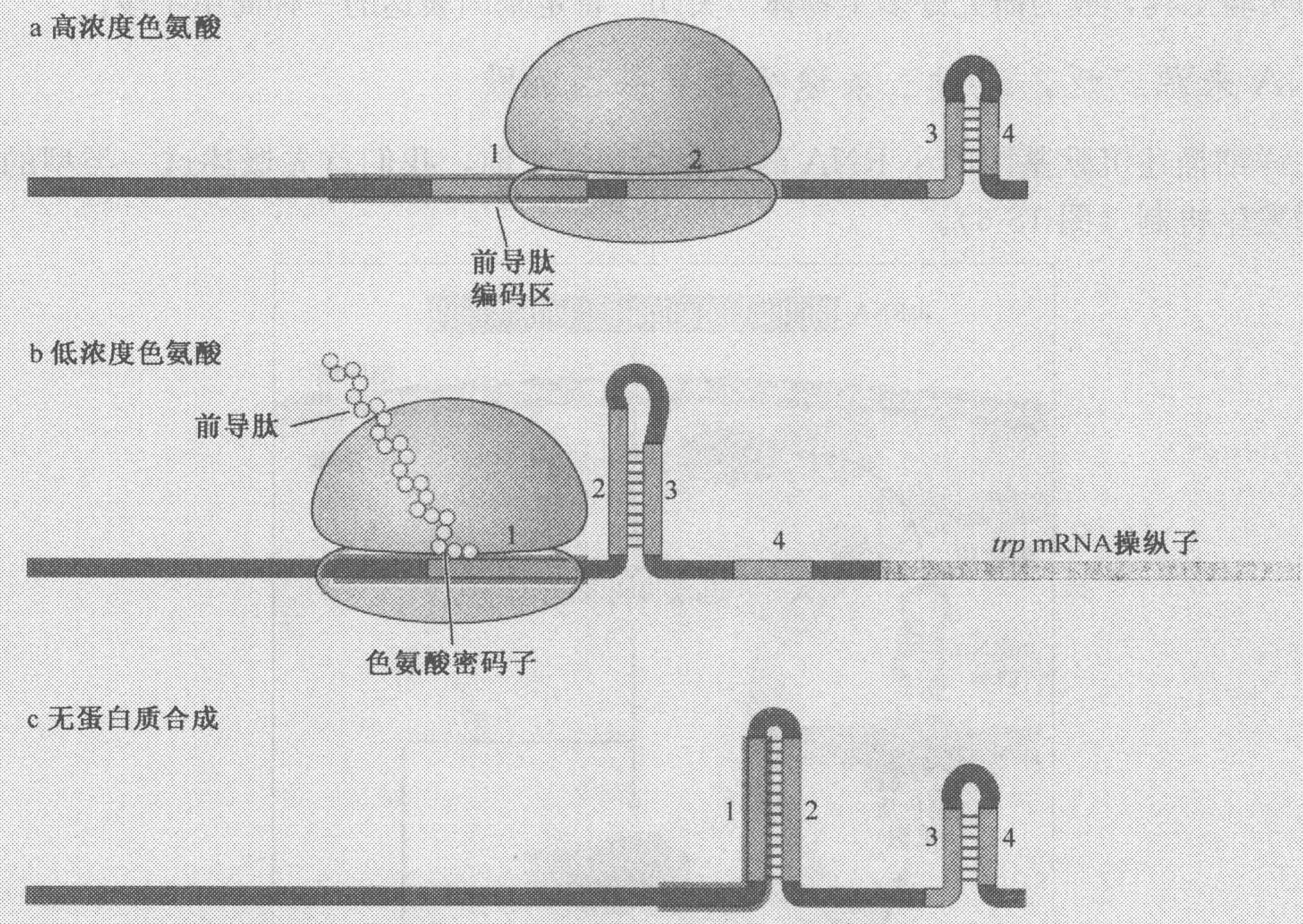
中的区域 3 和 4 序列而来），之后是 8 个 U 残基（图 12-12）。转录通常在这个终止子之后停止（并且我们可以设想，应该总是在此停止），产生一个长 139 个核苷酸的前导 RNA。这是色氨酸水平高的时候产生的 RNA 产物。

当细胞中色氨酸浓度低时，前导序列的三个结构特点使 RNA 聚合酶能够越过终止子。首先，前导 RNA 区域 1 和 2 之间形成的第二个“发夹“结构（除了终止子发夹之外）（框 18-1 图 2）。其次，区域 2 也与区域 3 互补；因此，在区域 2 和 3 之间又可以形成一个发夹结构，当这一发夹结构形成时，阻止了终止子发夹结构（区域 3 和 4 之间）的形成。第三，前导 RNA 含有一个编码 14 个氨基酸的短“先导”肽的可读框，这一可读框之前是一个很强的核糖体结合位置（框 18-1 图 2）。

编码先导肽的序列有一个显著的特点：两个色氨酸密码子排成一行。当色氨酸缺乏时，可利用的负载色氨酸 tRNA（charged trp tRNA）很少，核糖体移动到两个色氨酸密码子时它就会停顿下来。在这种情况下，色氨酸密码子周围的 RNA 位于核糖体里，并且不能形成发夹环。框 18-1 图 3 显示了这一结果。

一个核糖体停在色氨酸密码子上（图 3b）覆盖住了区域 1，剩下区域 2 可以自由和区域 3 配对；因此，终止子发夹（由区域 3 和区域 4 形成）不能形成，转录不会被衰减。但是，如果有充足的色氨酸（也有足够的负载 Trp tRNA）使核糖体继续通过色氨酸密码子，那么核糖体会阻碍序列 2，此时含有区域 3 和 4 的 RNA 就会形成。因此，终止子形成，转录衰减，*trp* 基因不能被转录。

Trp 操纵子受抑制作用和衰减作用的控制，提供两步反应来应答更加严酷的色氨酸缺乏。但是单独的衰减作用即可提供精巧的调节：其他的氨基酸操纵子如 *leu* 和 *his* 完全依赖衰减作用来调节。在亮氨酸操纵子中，前导肽含有 4 个相邻的亮氨酸密码子，而在组氨酸操纵子中前导肽有一排 7 个组氨酸密码子。



框 18-1 图 3　*trp* 衰减子处的转录终止。图中显示了待用色氨酸如何在 *trp* 操纵子的衰减子处控制转录终止。蓝框显示的是先导肽的编码区。a. 色氨酸含量高的情况：序列 3 可以与序列 4 配对以形成转录终止发夹。b. 色氨酸含量低的情况：核糖体在相邻的色氨酸密码子处停顿，使序列 2 能与序列 3 自由配对，因此阻止了 4 终止发夹的形成。c. 无蛋白质合成：如果没有核糖体于 AUG 处开始翻译前导肽，序列 1 和序列 2 的配对形成发夹，阻止了 2-3 发夹的形成，并允许序列 3-4 发夹的形成。Trp 酶不表达。

RNA 干扰是真核生物的主要调控机制

我们已讨论了原核生物中 RNA 分子是如何调控基因表达的，我们还讨论了一些真核基因转录的调控 RNA 元件的例子（第 17 章），但是这些调控功能都是通过结合调控蛋白发挥作用的（HIV 的 TAT 蛋白就是其中一例）。然而，现在已经清楚，RNA 在真核生物基因调控中具有更广泛且难以预料的作用。我们将在本节中阐述这些调控的新模式。

真核生物中存在多种类型的很短的 RNA 分子，它们能够抑制或沉默那些与其具有同源序列的基因的表达，这种调控机制被称为 RNA 干扰（RNA inteference，RNAi）。它具有多种不同的表现方式：有时抑制 mRNA 翻译，在另外情况下降解 mRNA，甚至还能沉默指导 mRNA 转录的启动子。如下所述，这些小 RNA 分子是由较长的、不同来源的双链 RNA（dsRNA），通过特异性酶酶切而产生的。

至今尚不清楚此类调控 RNA 的范围究竟有多大，靶基因沉默的具体机制也只是刚露端倪。但是已经清楚，这些 RNA 在从发育调节到保护机体抵御细菌感染过程中都发挥着重要的作用。在秀丽新小杆线虫（*Caenorhabditis elegans*）和拟南芥（*Arabidopsis*）中，RNA 如何调控发育过程已被研究得较为清楚（第 22 章）。另外，RNA 干扰技术已经成为

有力的实验工具，成为研究好多生物体“关闭”特定基因表达的一种简单手段。

小 RNA 来源广泛，通过三条途径指导基因沉默

在详细描述沉默基因的小 RNA 的来源和功能之前，我们首先概述这一类型的基因沉默的发生机制（图 18-6）。

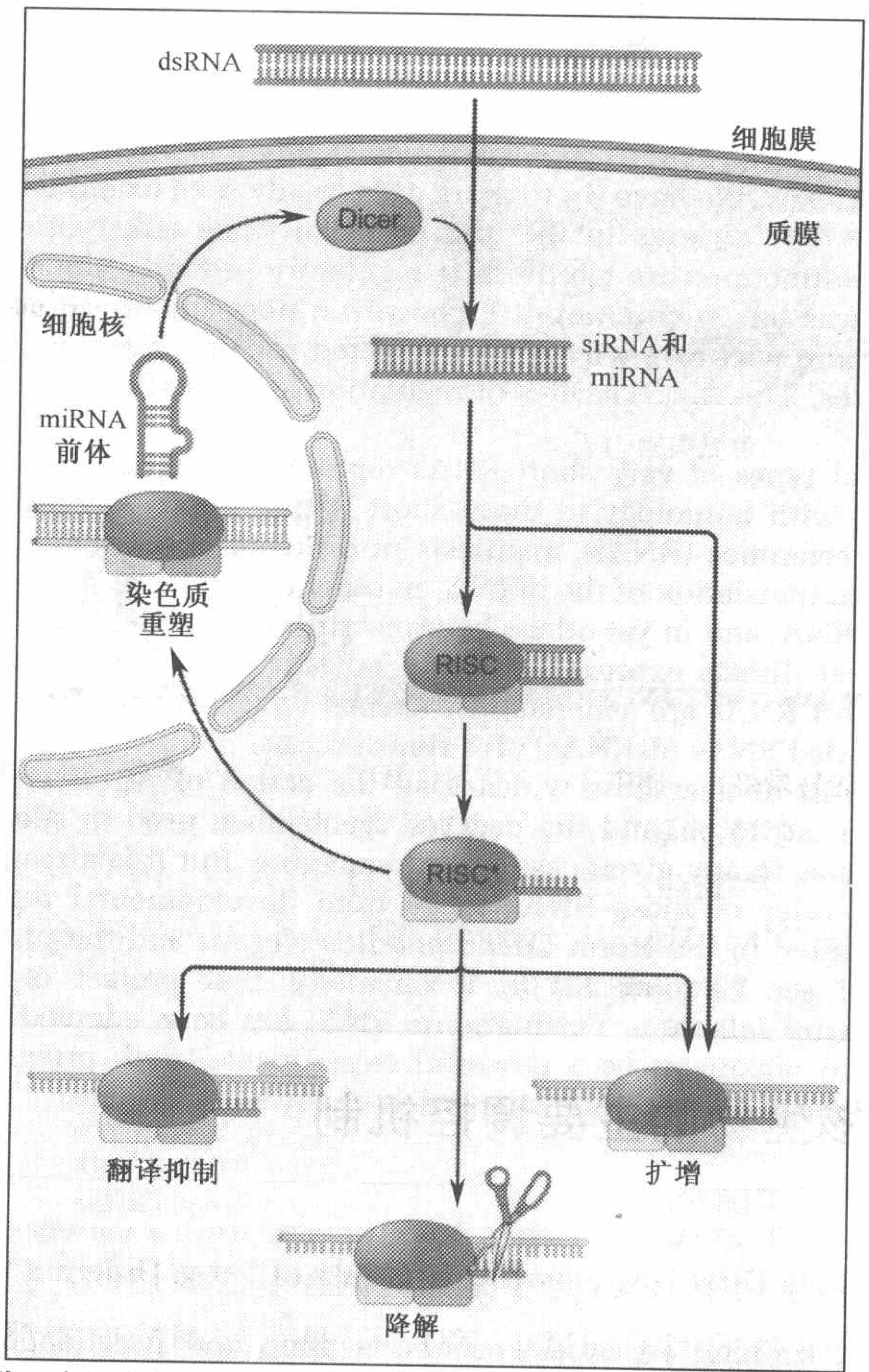


图 18-6 RNAi 沉默。当细胞产生或导入与某个基因同源的 dsRNA 时，RNAi 就会关闭这个基因的表达。这一效应与 Dicer 酶加工 dsRNA 成 siRNA 和 miRNA 的过程有关。另一个仅参与 miRNA 加工的酶（Drosha）在此没有显示，而将在后面描述。siRNA 和 miRNA 以三种方式指导一个称为 RISC（RNA-induced silencing complex，RNA 介导的沉默复合物）的复合物来抑制基因：攻击并降解与该 siRNA 同源的 mRNA；干扰这些 mRNA 的翻译；引导染色质修饰酶到达启动子来指导那些 mRNA 的表达。（自 Hannon G. J. 2002. *Nature* 418：244-251，Fig. 5. ©Macmillan. 经许可修改）

小 RNA 是以来源命名的。源于活体的双链 RNA 前体的或人工制备的小 RNA 被称为小分子干扰 RNA（small interfering RNA，siRNA）。另一类调控 RNA 是微 RNA（microRNA，miRNA）。后者是由细胞内基因编码的 RNA 前体衍生而来。这些miRNA 在细胞中具有特定的调控功能。

siRNA 和 miRNA 都是由长 RNA 分子经 Dicer 酶消化而来的。Dicer 酶是一种 RNase-Ⅲ类酶（图 5-24）。它能识别并消化长双链 RNA 或者 miRNA 前体形成的柄-环结构（见下文）。

siRNA 和 miRNA 产物一般长度为 21～23 个核苷酸，其产生过程见图 18-6 中的第一步。

这些小 RNA 通过三种途径来抑制同源靶基因的表达：第一种是触发靶基因编码的 mRNA 的降解；第二种是抑制 mRNA 的翻译；第三种是对靶基因所在染色质进行修饰而沉默其转录。显而易见，在所有的例子中，无论采用哪种途径，都需要很多相似的“机器”。这个“机器”包括 RNA 诱导的沉默复合物（RNA-induced silencing complex，RISC）。除了 siRNA 或 miRNA 之外，一个 RISC 还含有许多蛋白质，包括 Argonaute 蛋白家族的一个成员。

无论是 siRNA 或者 miRNA，都需要经过变性而形成一条向导 RNA（guide RNA）和一条搭载 RNA（passenger RNA）。向导 RNA 使 RISC 具有特异性，而搭载 RNA 通常被丢弃。所生成的复合物（即成熟的 RISC）被引导到含有与向导 RNA 序列互补的靶 RNA 上，靶 RNA 被降解，或其翻译被抑制。通常，降解或抑制依赖于向导 RNA 与靶 RNA 的配对程度：假如序列高度配对，靶 RNA 将被降解；如果配对不全（如一些碱基对间有错配），翻译被抑制更为常见。在靶 RNA 被降解的情况中，Argonaute 的催化亚基负责开始剪切 mRNA。因此，Argonaute 经常被称做“切片机”（slicer），mRNA 的剪切又称“切片”。

RISC 也可以被引导到细胞核中。它通过调动染色质修饰蛋白，汇集在与向导 RNA 互补的靶基因启动子上（图 18-6 左侧所示）。该修饰能够导致转录沉默（第 17 章）。例如，要沉默裂殖酵母（*Schizosaccharomyces pombe*）的着丝粒区域的转录，就需要 RNAi“机器”。在该例子中，着丝粒区域被认为能转录 RNA（第 7 章），这些 RNA 能够同来自于同一区域的其他 RNA 杂交形成 dsRNA，而该 dsRNA 被 Dicer 识别并被切割，产生的 siRNA 能够引导 RNAi 机器到中心粒区域。

RNAi 沉默的另外一个值得注意的特性是极其高效。因此，只需要很少量的dsDNA 就足以引发靶基因表达几乎彻底关闭。RNA 依赖 RNA 聚合酶（RNA-dependent RNA polymerase，RdRP）是提高 RNAi 效率的重要因子。很多情况下，如裂殖酵母着丝粒的沉默，RdRP 还是 RNAi 所必需的另一个酶。被原先的 siRNA 聚集到 mRNA 上后，RdRP 通过合成 dsRNA，来放大抑制信号（图 18-6 右侧所示）。这一反馈过程能够产生大量的 siRNA。尽管在哺乳动物细胞中 RdRP 的存在目前还肯定，但是 RNAi 高效性仍然可能来自于切割的催化性，即每个 RISC 可以切割好多个 mRNA。

因此，虽然在本章的第一部分我们已看到许多关于小分子 RNA 如何通过反式作用（*in trans*）调控细菌基因表达的例子，但是与真核生物中这些 RNA 的产生和作用的机制完全不同。

miRNA 分子的合成与功能

miRNA 的特征性结构——自身和靶基因识别

我们知道，miRNA 是一类通过 RNAi 途径来使基因表达沉默的调控 RNA。事实上，从调控 RNA 的产生和功能方面来说，miRNA 可能是被研究最彻底的一种。如前

所述，miRNA 是由基因组中较长的转录物片段编码的。它们这种特征性结构，有助于识别它们，并预测可能受它们调节的靶基因。

典型的功能性 miRNA 长 21 或 22 个核苷酸，其长度变化范围也可以从 19～25 个核苷酸。这种小 RNA 是由一个携带有发夹形的二级结构的长 RNA 转录物（称为初级 miRNA，pri-miRNA），经过两步剪切反应而产生的。第一步剪切反应是释放柄-环，称为前体 miRNA（pre-miRNA）；第二步由 pre-miRNA 的产生成熟 miRNA。最先被鉴别，并很好阐述其功能的 miRNA 是 *let-7*。*let-7* 调控线虫 *C. elegans* 从幼虫到成虫转变的发育过程（第 22 章）。图 18-7 显示的是 *let-7* 的 pre-miRNA 和一些其他的天然存在的 miRNA 的结构。

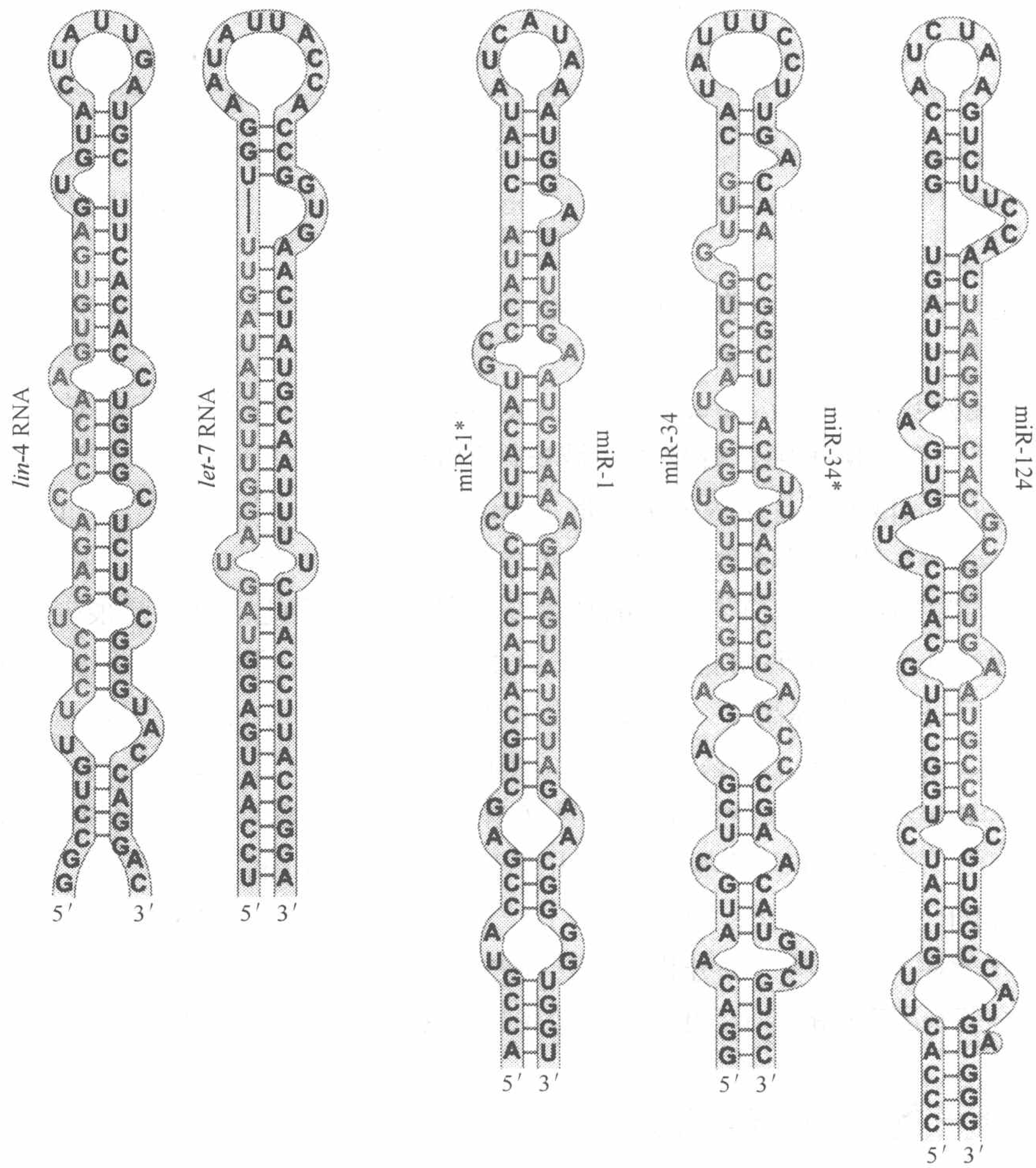


图 18-7 一些 pre-miRNA 的结构（加工成成熟的 miRNA 之前） miRNA 序列标为红色。在一些例子中，柄-环的两个“臂”都能产生有功能的 miRNA。在这些例子中，第二个 miRNA 用蓝色显示——譬如，miR-1（红色）和 mi-Rl*（蓝色），以及 miR-34（红色）和 miR*（蓝色）。显示的 miRNA 均来自于线虫。*lin-4* 和 *let-7* 是以遗传学方法鉴定的；而那些 miR 是以生物信息学方法发现的。（自 Lim L. P. et. al. 2003. Genes dev. 17：991，Fig. 6. ©Cold Spring Harbor Laboratory Press，经许可改绘）

原先认为，pre-miRNA 的柄-环结构中，只有一条"臂"可能是调控 miRNA。然而，从最近大量的例子中发现，这种结构的两条"臂"都产生功能性 miRNA，每一条都有它自己的一组靶基因（图 18-7 中用红色和蓝色分别标出两种 miRNA）。现在看来，两条"臂"均能产生 miRNA 的现象是很普遍的。pre-miRNA 可由被转录物的任何一个区域所编码：即它们可能存在于转录物的编码区，前导区或者内含子中（图 18-8）。

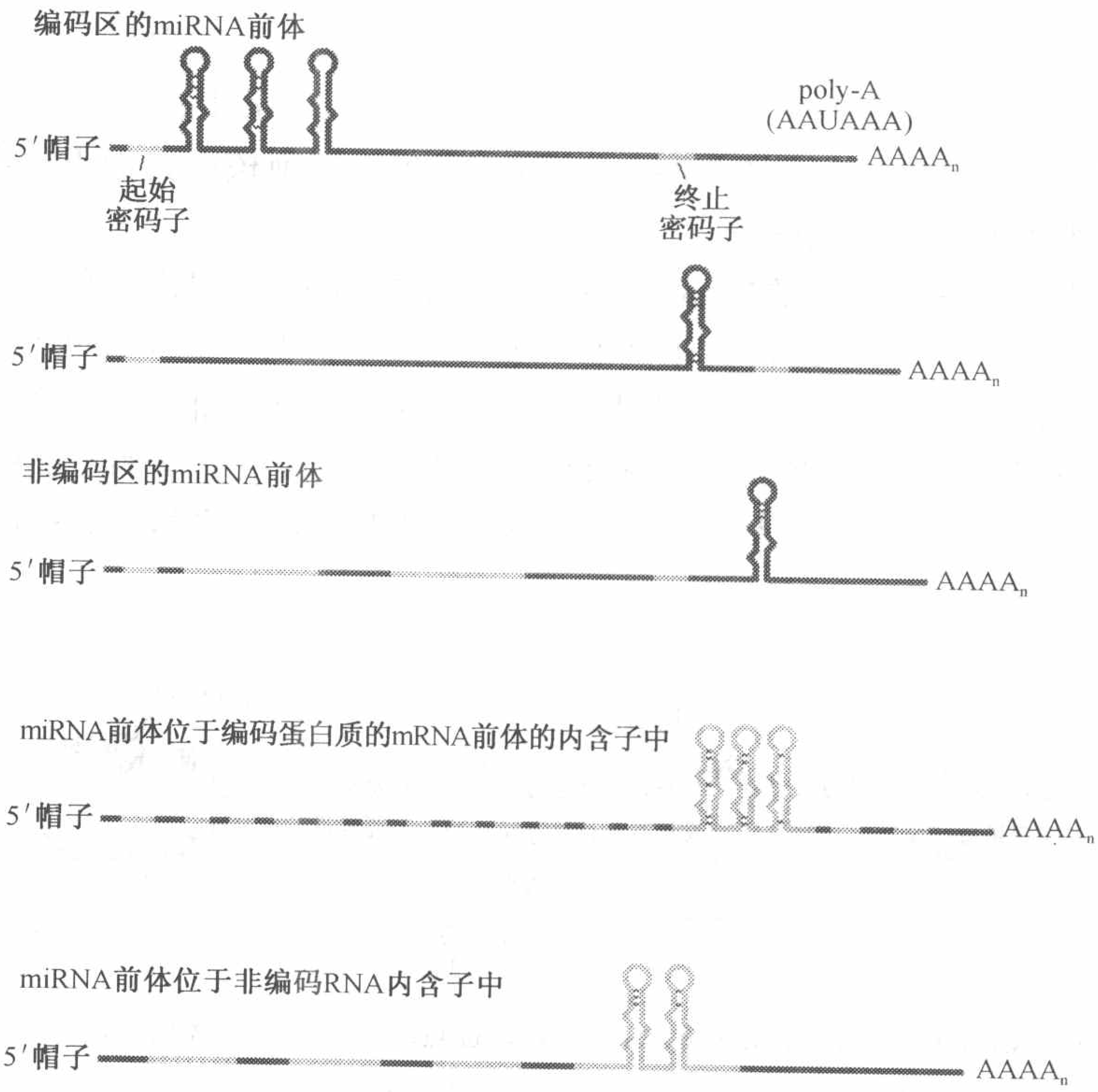


图 18-8　RNA 的内含子和外显子都可以为 miRNA 编码。内含子序列标以亮绿色。起始和终止密码子分别标以橙色和粉色。

由于 miRNA 的初级转录物（primary transcript of miRNA，pri-miRNA）具有特征性的二级结构，有可能通过计算 RNA 序列二级结构的折叠方式而进行预测。此外，在很多例子中，受调控的靶基因也同样可以被预测，因为沉默的发生依靠的是靶 RNA 和成熟 miRNA 之间的序列互补。miRNA 和靶 RNA 的碱基配对是通过"种子残基"（seed residues）的相互作用启动的。这些"种子残基"通常是指 22 个核苷酸 miRNA 中的第 2 到第 9 个核苷酸之间的序列。该序列与靶基因高度互补，所以这也是识别可能的靶基因最有效的区域。当然，要确认 miRNA 的存在，需要在细胞中被实际检测出来（如以 Northern 印迹），其存在能够影响靶 mRNA 的表达。

初级转录物经过两步切割反应产生 miRNA，需要两种不同的 RNase 酶来催化。一种是 Dicer，我们已经介绍过，几乎所有 RNAi 都离不开它。另外一种是 miRNA 加工

所特需的，这就是 Drosha。这两种酶的一个共同特点是它们对 RNA 的识别和切割基于底物的结构而不是特定的序列。我们现在着重讨论这些酶的功能。

活性 miRNA 是由两步核解过程产生的

两个特化的 RNA 切割酶是把含有柄-环结构的 pri-miRNA 初级转录物加工成成熟的 miRNA 所不可或缺的。第一个酶是 Drosha，是 RNaseⅢ酶家族的成员。Drosha 经过两次切割，将 RNA 的柄-环区（pre-miRNA）从 RNA 初级转录物（pri-miRNA）切下来。该酶与另一种必需的特异性蛋白亚基（在一些生物体中称为 Pasha，在另一些中则叫 DGCR8）一起形成微加工复合物（microprosessor complex）。由 Drosha 产生的 pre-miRNA 通常约长 65～70 个核苷酸。Drosha 存在于细胞核中，而 Drosha 催化的切割过程也常在细胞的这一区域进行。

pri-miRNA 中的碱基配对的柄部通常长约 33 个核苷酸（dsRNA 的 3 周螺旋），而且只含少量错配（图 18-9）。环部位于柄的"项"上（"顶环"），大小可变（相对较大，通常 10 个核苷酸左右）。对于加工反应来说，该环部的序列并非十分重要。重要的是，对于 Drosha 催化的加工反应来说，单链 RNA（ssRNA，没有显著的二级机构）一定要在柄-环的 5′和 3′两侧。在很大程度上，ssRNA-dsRNA 的连接（junction）决定了 Drosha 切割的特异性。

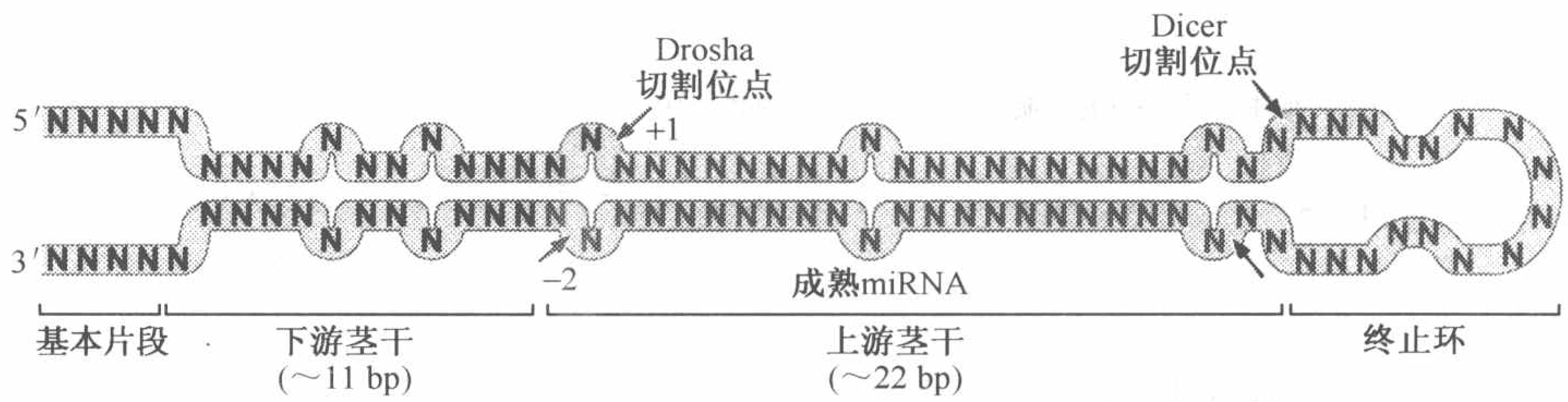


图 18-9 pri-RNA 结构概观，示 Dicer 和 Drosha 切割位点。红色标记的区域将被加工成成熟的 miRNA。注意基本片段必须为单链，以便 Drosha 复合物正确识别。

柄部可以被分为两个功能片段：一个约 11bp 的柄下部区和一个大约 22bp 的柄上部区（图 18-9）。Drosha 从位于 pri-miRNA 的柄下部和上部区之间的 ssRNA-dsRNA 连接上切割下 11bp（图 18-10）。因此这两次切割产生了长约 65 个核苷酸的 pre-mi RNA，由 22bp 双链 RNA（2 周螺旋）和"顶环"组成。RNaseⅢ家族的酶特异性识别 dsRNA 并将其切割，dsRNA 产物的 3′端带有一个核苷酸的"接头"（overhang）。这个 3′接头对于下一个酶对 RNA 分子的识别是很重要的。

Dicer 是 miRNA 生成的第二个 RNA 切割酶

Drosha 释放的 pre-miRNA 被输送到细胞质中，并在此接受由 Dicer 执行的第二步切割反应。像 Drosha 一样，Dicer 不是利用序列特异性，而是利用测量机制来选择切割位点。Dicer 的高分辨率结构向我们展示了这一切割过程可能是如何发生的。

Dicer 由三个结构域（module）构成（图 5-23）：两个 RNaseⅢ结构域和一个与 dsRNA 结合的结构域 PAZ。PAS 源自三个蛋白质名：Piwi、Argonaute 和 Zwille。

图 18-11a 为 Dicer 蛋白组织的模式图，并显示它是如何与 dsRNA 分子相互作用的。图 18-1b（右半部分）示 Dicer 的结构以及与底物 RNA 作用的模型。该蛋白质的整体结构像一个短柄斧（hatchet）。PAZ 结构域在手柄的底部，形成了 dsRNA 底物 3′端的结合处（binding pocket）。连接结构域（linker domain）形成了斧头的柄部，并含有一个带正电荷的 RNA 分子结合表面（binding surface）。顶部的“刀刃”区域包含了两个 RNase 结构域，为排列对称的二聚体。每一个 RNase 结构域带有一个活性位点，分别负责切割底物 RNA 两条链中的一条。因此，Dicer 可作用于任何 dsRNA，不管序列如何，并从底物 RNA 的一端切下生成一个长 22 个核苷酸的 mRNA 分子。PAZ 结构域像“标尺”那样（ruler-like），锚定在底物 RNA 的离 3′端长约 22 个核苷酸左右的酶活性位点处（图 18-11）。实际上，在不同生物中发现的 PAZ 结构域的大小与 Dicer 产物的大小有关。

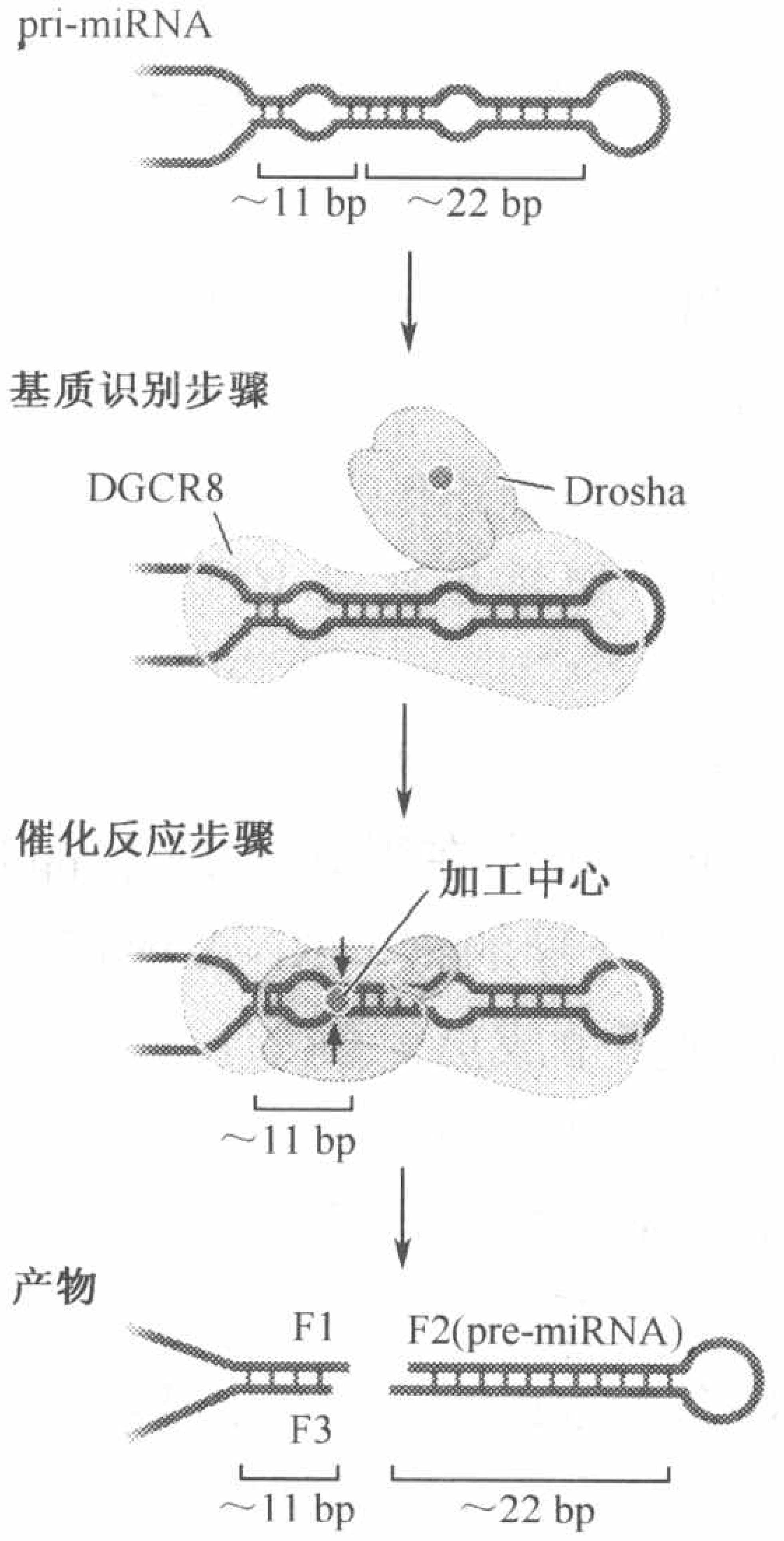


图 18-10　微加工复合物对 pri-RNA 的识别与切割。切割产生的三个片段分别标记为 F1、F2（pre-miRNA）和 F3。

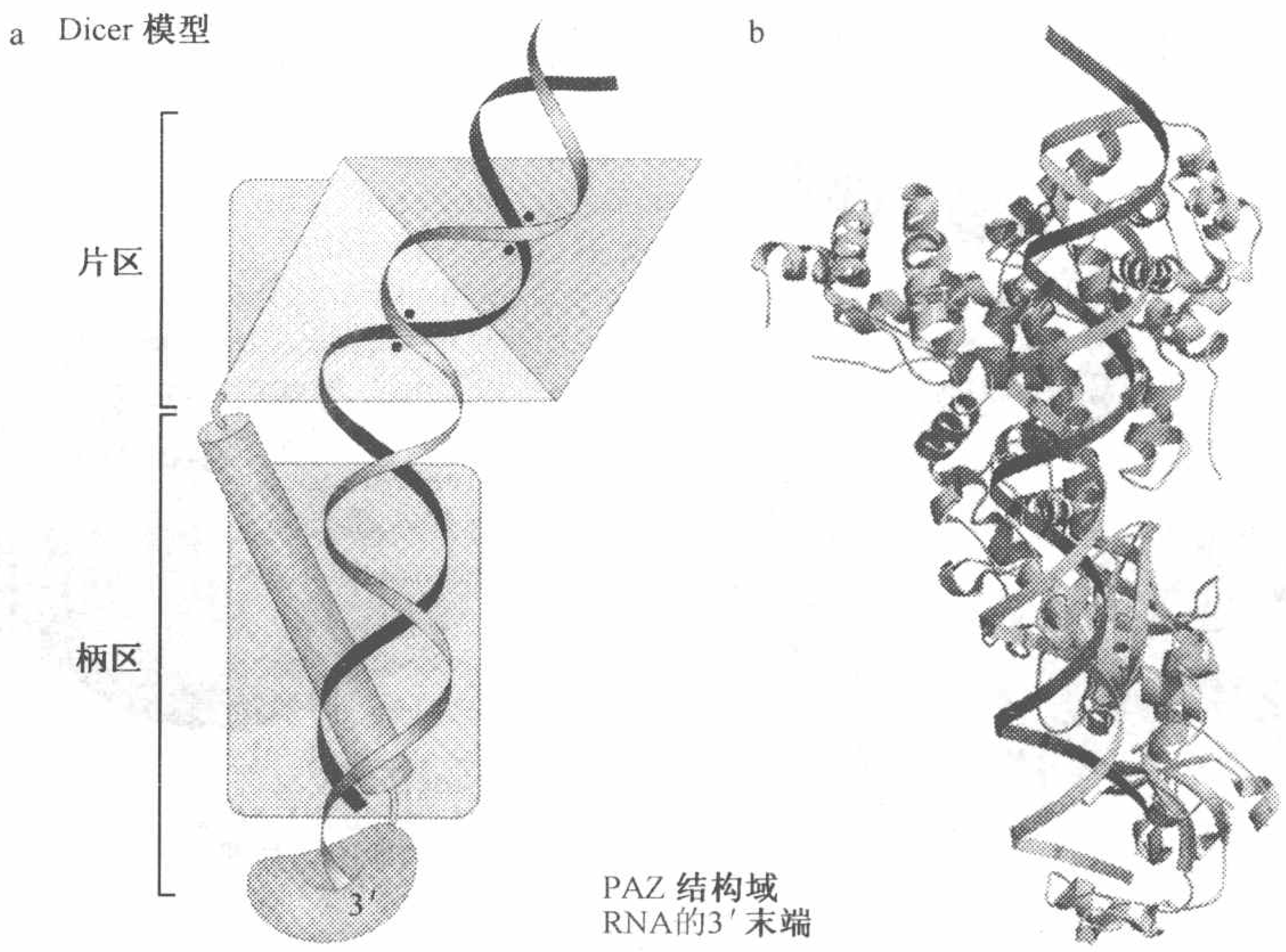


图 18-11　Dicer 的结构与组织。a. 图示 Dicer 的组织。b. Dicer 结合 dsRNA 的结构模式图，示其长度是如何被测定的。蛋白质示以灰色，红色箭头示核酸酶活性位点（相当于 a 中黑点表示的部分）。RNA 用绿色显示。这一结构仅显示了 RNaseⅢ和 PAZ 结构域。Dicer 酶也含有 ATP 酶和其他结构域。（b，MacRae I. J. et al. 2006. Science 311：195-198. PDB Code：2FFL；注意这是 RNA 的模式图，不是晶体结构。）图像以 MolScript、CobScript 和 Raster3-D 制备。

前导 RNA 链整合进 RISC，形成能沉默基因表达的成熟复合体

Drosha 和随后 Dicer 的作用产生了长 21～25bp 的 RNA 分子，能够指导基因表达的调控。活化形式的调控 miRNA 是指以单链形式整合到 RISC 上，在这一阶段称为“指导” RNA（guide RNA）。在这个复合物中，指导 RNA 聚集 RISC 到靶 RNA 上。人们还在争论的是：在复杂的真核生物的巨大基因组中，22 个核苷酸左右的长度是否足够，以满足通过 RNA-RNA 碱基配对来准确无误地识别靶基因。

RISC 是一个含有指导 miRNA 的多组分复合物。指导 RNA 与靶 mRNA 进行碱基配对，促进基因表达的沉默。这一调控复合物的中心组分是一个 Argonaut 蛋白。Argonaut 蛋白在很多情况下都是 RNA 剪切酶。对基因沉默机制理解得最为透彻的以 RISC 介导的靶 mRNA 剪切（或“切片”）。然而，很多生物中都含有 Argonaut 蛋白家族的多个成员。例如，在人体中就含有 8 种不同的 Argonaut，但不是所有的 Argonut 在整合到 RISC 中后都有沉默活性。含有其他 Argonaut 的 RISC 必须利用非切割依赖的机制来沉默基因表达，如翻译抑制。

下面将介绍活性 RISC 的产生和切割发生过程。首先，通过 Dicer 产生的短 dsRNA 整合进 RISC。dsRNA 然后变性产生一条指导链和一条搭载链。根据第一个模型，搭载链被 Argonaut 切割后就被排除出复合物；而根据另一个模型，搭载链的排除不需要先被切割，可能是被 RNA 解旋酶排除的。所生成的 RISC（称为成熟 RISC）连同它的单链指导 RNA，具有识别和切割靶 mRNA 的活性。

正如我们在 Dicer 中所观察到的那样，Argonaut 蛋白的结构为了解 RISC 如何识别和剪切靶 mRNA 的机制提供了框架（图 18-12）。跟 Dicer 一样，Argonaut 也含有 PAZ 结

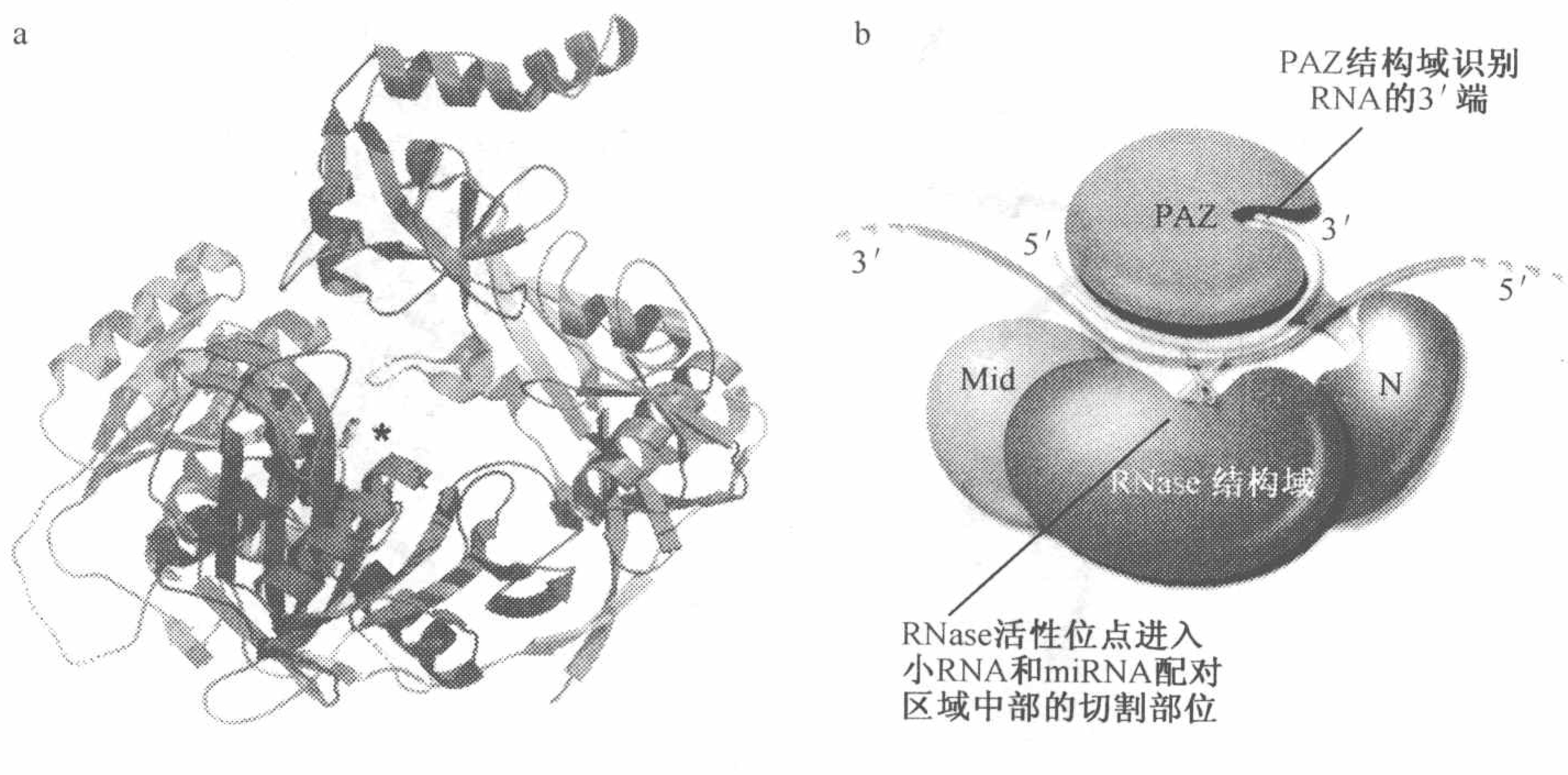


图 18-12 Argonaut 蛋白的结构，显示 RNA 结合区和 RNaseH 样核酸酶结构域。 a. Argonaut 蛋白的晶体结构。结构域如 b 中着色部分所示，蓝色区域显示蛋白质的氨基端，绿色为中间的结构域。B. Argonaut 结构域的模式图。箭头所指的 RNase 的活性位点，位于小 RNA 和 miRNA 配对区域中部的切割部位。（a，Song J. J. et al. 2004. Science 305：1434-1437，Fig. 4C. PDB Code：lu04. 图像以 MolScript、CobScript and Raster3-D 绘制。B，自 Song J. J. et al. 2004. Science 305：1434-1437，Fig. 4C. ©AAAS，经允许改绘）

构域和 RNase 结构域。PAZ 结构域能够特异地识别指导 RNA 的 3′端。指导 RNA 通过碱基配对结合到靶 RNA 上，这一复合体的构造恰好使得靶 RNA 落在 RNase 结构域的活性位置上，使靶基因链易于被剪切。剪切部位靠近“指导 RNA-靶 RNA”二聚体的中央，在指导 RNA5′端第 10 个和第 11 个核苷酸之间。

如前所述，在某些情况下，成熟的 RISC 不是剪切 mRNA 而是抑制它的翻译。这种翻译抑制的原理还在研究。但据观察显示，miRNA 有时能引导 mRNA 螯合在细胞质里的“加工体”（processing body，P-body）上，翻译便被抑制。miRNA 的结合也会导致靶 mRNA 的 ploy-A 尾巴的不稳定，抑制翻译的启动。

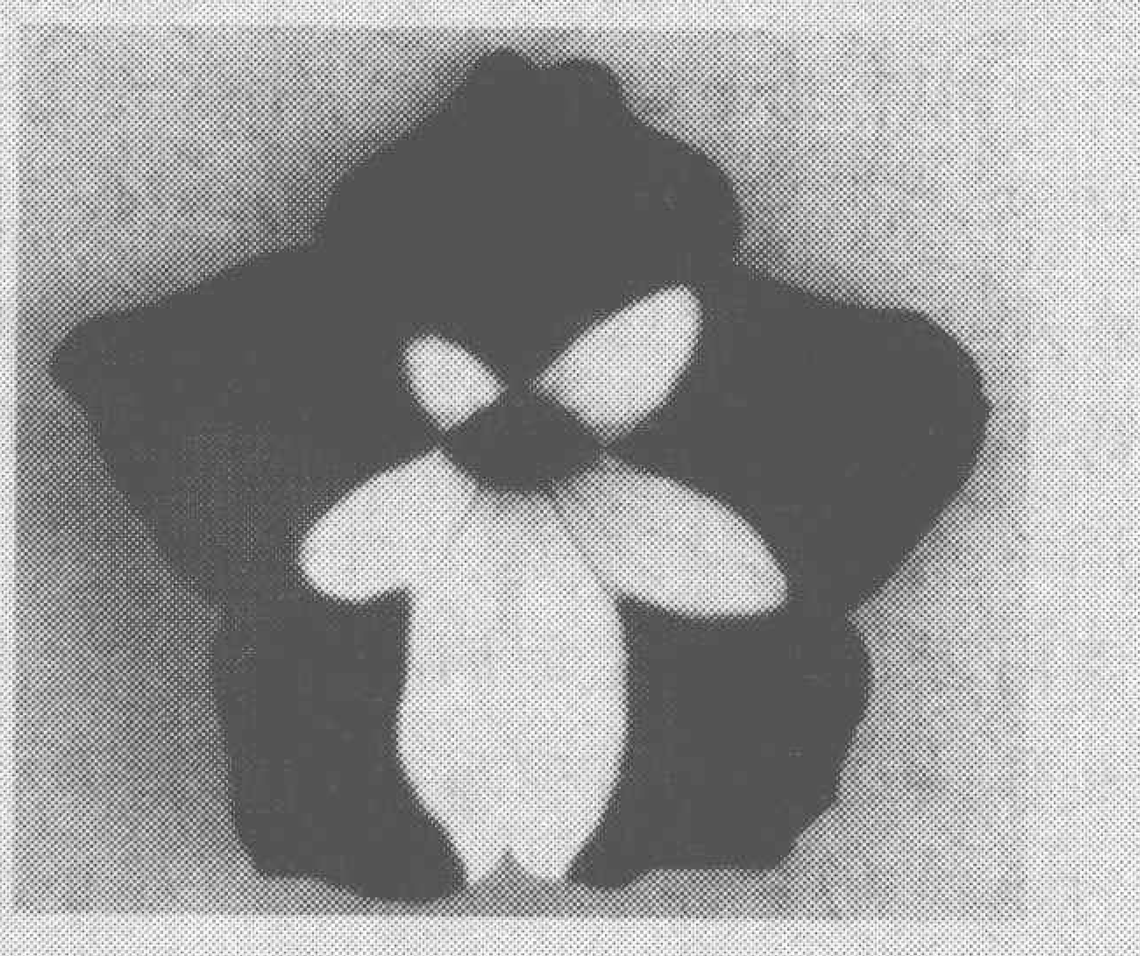
siRNA 是源于双链长 RNA 的调控 RNA

正如我们所看到的，形成活性 miRNA 的关键步骤包括：Drosha 剪切、Dicer 剪切，然后是指导 RNA 整合进 RISC。只有 miRNA 是由大的发卡前体加工形成的。与此相比，siRNA 的前体 RNA 是较长的双链 RNA。由于使用不同的起始底物，siRNA 的形成不需要 Drosha。但 Dicer 剪切还是需要的，以形成 21 或 22 个长度核苷酸的 RNA 适合于整合进 RISC。在植物中，甚至 miRNA 只需要 Dicer 就够了，尽管还不清楚它们是怎样跳过前面的 Drosha 步骤的。

一系列引人入胜的观察和实验使我们现在对真核生物的小调控 RNA（regulatory RNA）有了了解。这些努力开始于 20 世纪 80 年代末，当时试图在矮牵牛花中高表达色素基因，却得到了令人费解的结果（本来想使紫色加深，结果却得到白色的花）。下一个令人惊奇的发现，是线虫中的调控基因的产物居然是 miRNA。进一步的实验显示，将双链 RNA 导入线虫体内，沉默了互补基因的表达。这些故事将在框 18-2 中详述。

框 18-2　miRNA 和 RNAi 的历史

1989 年，在美国加州奥克兰（Oakland）Advanced Genetic Sciences 生物技术公司工作的 Richard Jorgensen，试图培养出比现有品系花色更深的紫色矮牵牛花。这一策略看似直截了当：他打算将一个强启动子控制的色素基因（编码查耳酮合成酶，chalsone synthetase）的额外拷贝导入植株之中。这些植株将合成更多的查耳酮合成酶，从而使花的紫色加深。他真正得到的植株却是浅淡不一的矮牵牛花，其中很多紫白相间，有的甚至全为白色（框 18-2 图 1）。



框 18-2 图 1　矮牵牛花。色素基因查耳酮合成酶过量表达效应的一个例子，正常情况下则是单一紫色的矮牵牛花。（Arizona 大学的 Richard A. Jorgensen 惠赠）

失望之余，这一结果还是令人惊奇的。在试图理解这一结果的过程中，Jorgensen 发现了这一现象，称为共抑制（cosupression）的许多特点（转入的基因和机体内源的基因表达都被抑制了）。转基因的表达水平超高，查耳酮合成酶的

水平越低；无论是因为转基因有很多拷贝，还是因为转基因的启动子很强，这种情况都存在。研究还观察到一些植株色彩斑驳（色素形成的杂色模式），并且可在同一植株的不同花朵上出现。这种模式有时是可以遗传的，但在其他情况下明显是随机可变的。这些观察结果使 Jorgensen 和其他研究人员（尤其是同样正在研究这一现象的 Marjori Matzke）想到这是一种表观遗传（epigenetic）现象。

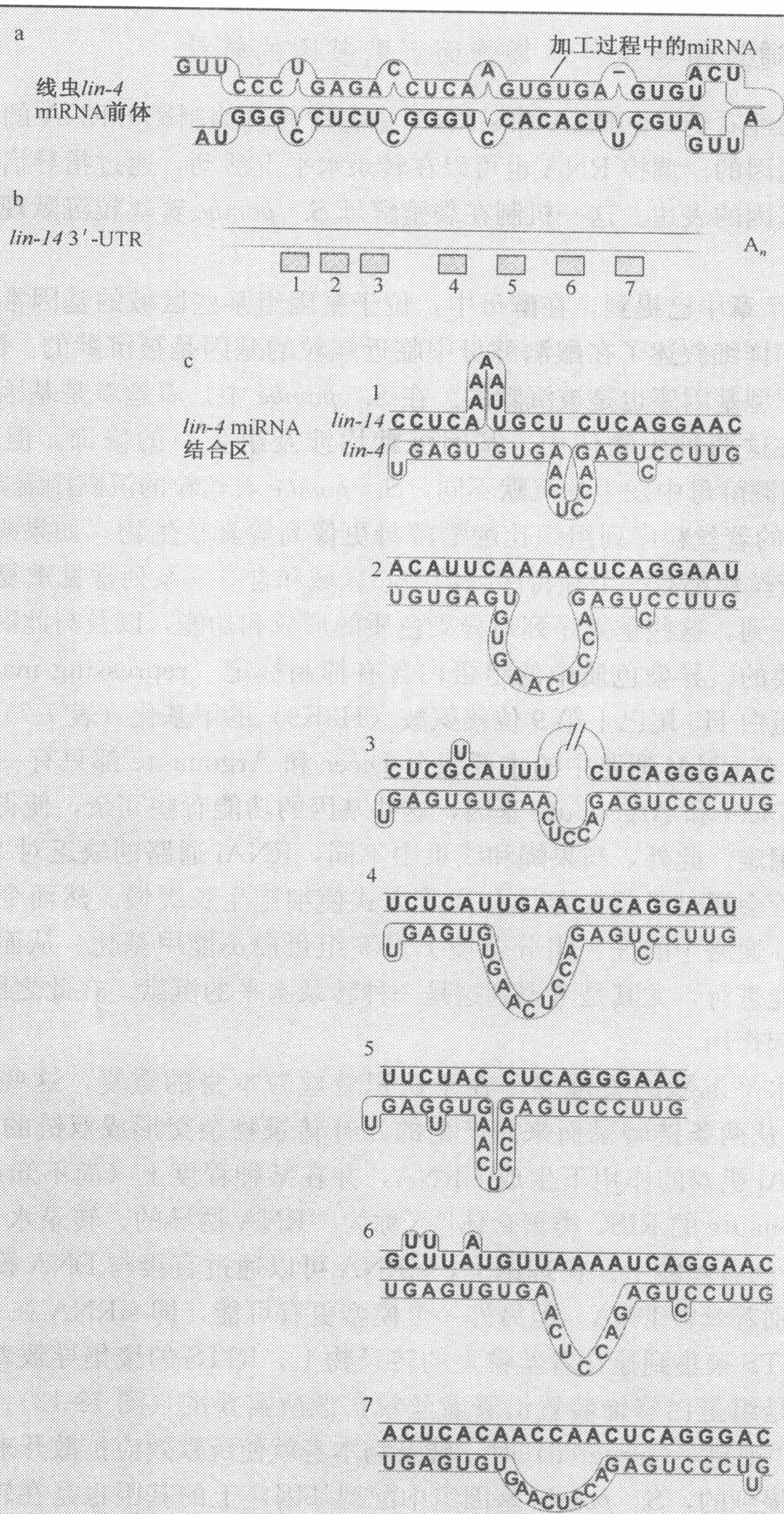
其他的研究者试图使植物获得抗病毒侵染的特性。一种途径是在植株中过量表达一种常见的病毒复制因子显性失活（dominant-negative）的衍生物，这种蛋白质被认为可用来阻断任何使用共同的复制机器的侵染病毒的复制。尽管这种源自病毒的显性失活的病毒产物阻断了这个马铃薯病毒的复制，令人惊奇的是，对这种病毒的作用的特异性被严格地限制，而且还证明这一过程甚至并不需要蛋白质本身，仅仅 RNA 就够了。

同时，其他的研究者使用反义 RNA（antisense RNA）在线虫中敲除了 *par-1* 基因的表达。他们本想证明这个基因与特定的发育性状相关。离体合成的反义 RNA 注入正在发育中的线虫，所诱导的性状符合原先的预期即 *par-1* 没有表达。但是，同义 RNA（sense RNA）居然也显示了同样的效应。在这个实验中，同义 RNA 仅是作为一个阴性对照，它并未被考虑到会对基因表达起任何作用。与 *par-1* 基因无关的 RNA 不起任何作用。

对他们实验中依赖 RNA 的基因表达的解释使 Andrew Fire 和 Craig Mello 赢得了 2006 年诺贝尔生理学或医学奖。他们证明事实上既不是正义 RNA 也不是反义 RNA，而是这两种的混合产物 dsRNA 沉默了基因。后来发现，以反义和正义 RNA 制备的 RNA 被少量的互补链相互污染，所产生的双链分子使基因沉默。特意制备的 dsRNA 显示能非常有效地抑制靶基因的表达。至此，RNAi 的现象被发现并在 1998 年发表。

好几个实验室的工作很快加深了对这一机制的了解。首先，dsRNA 被发现可以引起从果蝇细胞中提取的同源 mRNA 的降解，这一实验鉴定了 RISC。指导 RISC 结合到靶基因上的 siRNA 于 1999 年在植物中被鉴定并被报道。制造 siRNA 的核酸酶 Dicer 是在 2001 年被描述的。这一通路中的最后一种重要组成，Slicer 是在 2005 年被鉴定的，Argonaute 的晶体结构揭示这个蛋白质是一种 RNase。

除了为生成 siRNA 所需要外，Dicer 还是发育过程中 miRNA 起作用所需要的。1993 年，Victor Ambros 和 Gary Ruvkun 分别描述了第一个 miRNA 和它的靶基因。在那个时候，这一发现被认为是既巧妙又古怪的。*lin-4* 基因编码作用于靶基因 *lin-14* 的小 RNA，这种小 RNA 通过 miRNA 和靶基因 5′UTR 区之间的互补序列起作用（框 18-2 图 2）。随后，另外一些 miRNA 在线虫中被发现，一些与动物和植物中相似的基因同源，说明这种调控机制可能更普遍地存在。当 siRNA 和 miRNA 通过相同的途径起作用这一点明确后，所有的事情都被联系在一起了。关于基因调控的小 RNA 世界的画面也逐渐浮现出来，一些小 RNA 来自外源，另外一些是发育过程中基因调控程序的内部装置。这个领域的发展非常迅速，从这些现象被发现和解释，到摘取诺贝尔奖的桂冠，到写入教科书的章节，仅用了 15 年时间。这种快速进程应归咎于应用多种方法（遗传学、生物化学、结构研究和生物信息学）对多个物种（酵母、植物和线虫）的研究。



框 18-2 图 2 结合到靶基因 *lin-14* 的 3′UTR 区内的 microRNA*lin-4*。a. 被 Dicer 剪切前的 *lin-4* pre-miRNA 前体，miRNA 的序列标以蓝色。b. *lin-4* 的 3′UTR 中的 7 个序列，c 示其可与 *lin-14* mRNA 在不同程度上进行碱基配对。（据 Ha I. 1996. Genes Dev. 10：3041-3050，Fig. 1©Cold Spring Harbor Laboratory Press 修改）

小 RNA 可以通过指导染色质修饰而沉默基因的转录

我们现在已经看到，miRNA 和 siRNA 是怎样通过抑制靶 mRNA 的翻译或导致它们降解而沉默基因的。调控 RNA 也可以在转录水平上活动，通过指导启动子的组蛋白修饰而关闭靶基因的表达。这一机制在裂殖酵母 *S. pombe* 着丝粒沉默现象中得到了深入的研究。

我们在第 17 章中已提到，在酵母中，位于基因组某些区域的基因常常被沉默。在那一章里，我们详细叙述了在酿酒酵母中临近端粒的基因是被沉默的。酿酒酵母和 *S. pombe* 酵母的配型基因座也是被沉默的。在 *S. pombe* 中，着丝粒是基因组中另一个被沉默的区域。在这两种生物体中，基因沉默均涉及组蛋白的修饰。但是，与不需要 RNAi 参与的酿酒酵母中发生的沉默不同，*S. pombe* 着丝粒的沉默则需要 RNAi 通路。

S. pombe 的着丝粒序列组织比酿酒酵母更像高等真核生物（如果蝇和人类）（图 7-8）。它们的着丝粒都有一个独特序列的中心区域和含有一系列常见重复序列的两侧区域。我们将要看到，这些重复序列对异染色质的形成和功能，以及与此区域相关的转录沉默都是很重要的。异染色质中的组蛋白含有抑制标记（repressing marker）：低水平的乙酰化、组蛋白 H3 尾巴上第 9 位赖氨酸（H3K9）的甲基化（表 7-7）。

S. pombe 在 RNAi 通路中的主要组分 Dicer 和 Argonaute 都只有一个基因。高等生物具有多个 Dicer 和 Argonaute 基因，这些基因的功能有些冗余，使得 RNAi 通路的遗传操作更加困难。此外，与果蝇和线虫中不同，RNAi 通路的缺乏对 *S. pombe* 并不是致命的，尽管会通过干扰染色体分离等方式使细胞生长缓慢。然而令人吃惊的发现是，失去 RNAi 通路中的任一组分会使 H3K9 组蛋白不能甲基化、从而导致着丝粒的基因沉默也不能进行，尤其是考虑到这是一种转录水平的沉默。在此之前，RNAi 被认为仅在转录后起作用。

认识转录水平沉默的关键似乎在于认识着丝粒本身的重复，这些重复序列是由 RNA 聚合酶Ⅱ从两条链转录而来，产生的互补转录物杂交形成双链的 dsRNA。这种 dsRNA 在 RNAi 机器的作用下生成 siRNA，并在某种程度上（尚不知道如何起作用）指导“含 Argonaute 的 RISC 类聚合体”（称为“RNA 诱导的、转录水平沉默的 RITS 聚合体”）结合到着丝粒上。在理论上，siRNA 可以通过直接与 DNA 模板的序列特异性配对，来识别着丝粒 DNA。但另外一个模型更有可能，即 siRNA 在 RNA 聚合酶Ⅱ作用下，将 RITS 聚集到拴在着丝粒上的转录物上。RITS 的聚集导致着丝粒转录物的沉默，而这又是组蛋白修饰装置沿着着丝粒扩散所需要的（图 18-13）。这样，那些转录物被 RNAi “锁定”（targeted）时，转录物本身就使沉默效应扩散开来。

如上文所提到的，*S. pombe* 基因组中配型基因座上的基因也是在转录水平上被沉默的，在 RNAi 缺陷型突变株中，这种沉默的现象依然存在。在这种情况下，RNAi 据认为能照样作用，但只是在沉默的起始阶段起作用。沉默一旦建立，在维持阶段就不需要 RNAi 了。另外，如同在酿酒酵母中一样（第 17 章），基于蛋白质的机制将维持基因抑制的状态。据认为，在从果蝇到植物的其他生物中，RNAi 在异染色质沉默中也起到一定作用。如下文所述，沉默不需要的转座子转录物也是 RNA 介导的。

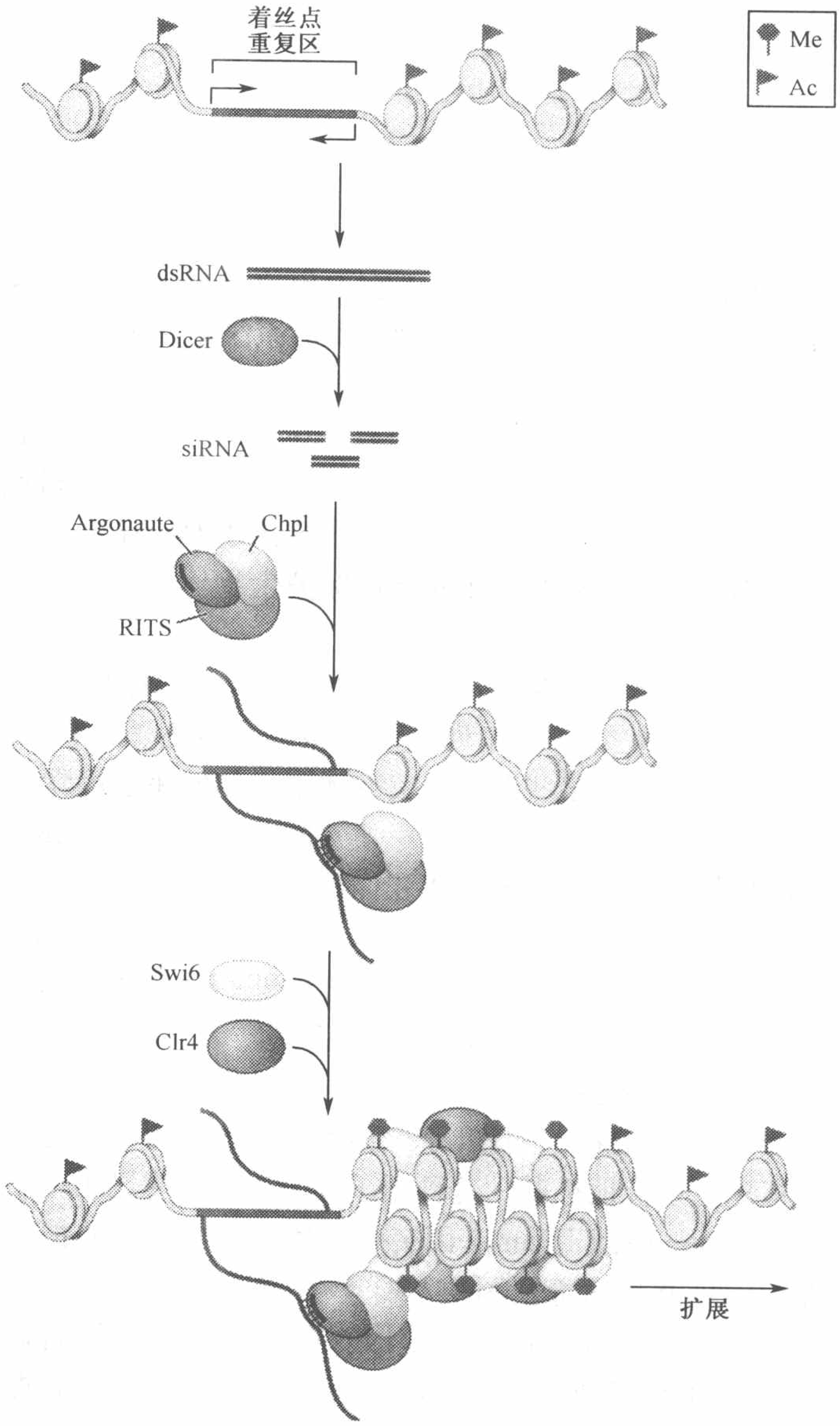


图 18-13 RITS 聚集和着丝粒沉默的一个模型。图的上部显示的是在 *S. pombe* 的着丝粒重复序列周围的核小体。这些重复序列是由 RNA 聚合酶Ⅱ从两条链上转录的，生成的 dsRNA 成为 Dicer 的底物。生成的这些 siRNA 被装载到含 Argonaute 的聚合体 RITS 上。如图的中部所示，含有 siRNA 的 RITS 聚合体被聚集到 PolⅡ拴住的转录物上，这些转录物是由着丝粒的重复序列通过 siRNA 和转录物的互补持续转录生成的。这个复合体随后聚集核小体修饰因子（Clr4 和 Swi6），这一修饰是通过标上 H3K9 沉默标记来实现的。RITS 的另一个亚基 Chpl 含有一个铬结构域（chromodomain）（图 7-40），通过与甲基化的核小体结合而有可能稳定 RITS 的结合。尽管未在图中显示，由 Argonaute（在 RITS 中）介导的转录物“切片”（slicing）生成的是 RdRP 的底物 RNA，合成一条互补链后，进一步生成 Dicer 的底物。这一过程是核小体修饰的扩散必不可少的，这个区域就这样被沉默。（Martienssen R. and Moazed D. 2007。Epigenetics（ed. Allis et al），p. 157，Fig. 4. ©Cold Spring Harbor Laboratory Press，经许可重绘）。

RNAi 的进化和应用

RNAi 像免疫系统那样进化吗?

RNAi 机器在真核生物中即使不是无处不在，但也是普遍存在。像我们所提到的，在酿酒酵母中就没有 RNAi 机器。然而，这种基本的系统据认为存在于所有真核生物的最近的共同的始祖先中，但随后在某些世系中丢失了。

但是，什么是 RNAi 的生物学功能呢？当有 miRNA，当然，就需要 RNAi 机器来制造和利用那些“调控者（regulator）”，但在一些生物中有 RNAi 机制却没有 miRNA（如 *S. pombe*）。事实上，miRNA 据认为在进化中受惠于 RNAi 机器，而非是 RNAi 存在的原因。RNAi 机器的一个古老功能有可能是保护生物体免受转座子和病毒的侵害，这一功能现在依然存在。

转座子在所有的真核生物中都存在，在某些情况下，在基因组中占据相当一部分。以人类为例，我们的基因组中 45%的序列曾是转座子。转座子常常在转录水平上被沉默并被组装进异染色质。然而在某些 RNAi 突变体中，与转座子沉默有关的组蛋白修饰丢失了。此外，在植物和线虫中，已鉴定了一些和转座子相应的 siRNA。在一些植物或线虫中，RNAi 的丢失激活了转座子，使它们跳跃，产生高水平的自发突变。然而，并没有那么多的转座子像我们知道的能被激活来产生 SiRNA 的。这似乎反映了在上面提到的 *S. pombe* 配型基因沉默的情形：RNAi 对于某些转座子的沉默启动是必需的，但启动之后的沉默则是自我维持，不再需要 siRNA。

RNAi 保护植物免受病毒侵染的作用非常普遍。事实上，这一作用早在 RNAi 机制被发现之前就已被认识。当植株的一片叶子感染了一种病毒，一种能够沉默这种病毒复制的因子就在整个植株上系统地扩散。这种因子不能保护原先已被侵染的叶子，但阻止了侵染的扩散。在 Argonaute 或 Dicer 基因突变的植株中，侵染的扩散没有受阻，病毒复制更快。这种保护信号就是病毒基因组本身产生的 siRNA。病毒也会报复：它们常携带一些保护病毒免受宿主 RNAi 作用的基因。一个例子是马铃薯病毒 Y 的 HcPro，它能降低马铃薯 siRNA 的产量或减弱 siRNA 的稳定性。其他一些病毒产物作用于防护系统的其他步骤，包括 siRNA 的系统扩散。

框 18-3 描述 RNAi 和人类疾病的关系。

框 18-3 RNAi 与人类疾病

癌症

在癌症中常见许多 miRNA 水平的降低。这种降低被认为揭示了这些 miRNA 通常具有抑制肿瘤的作用。除去这一常见趋势，其他一些特殊的 miRNA 在某些癌症中表达上调。同癌症相关的编码蛋白基因相似，这些 miRNA 被称为抑癌基因（如果失去它们将促发癌症）或致癌基因（如果它们的表达升高将导致癌症）。它们的靶基因常与细胞周期活动（细胞增殖）和细胞凋亡相关。

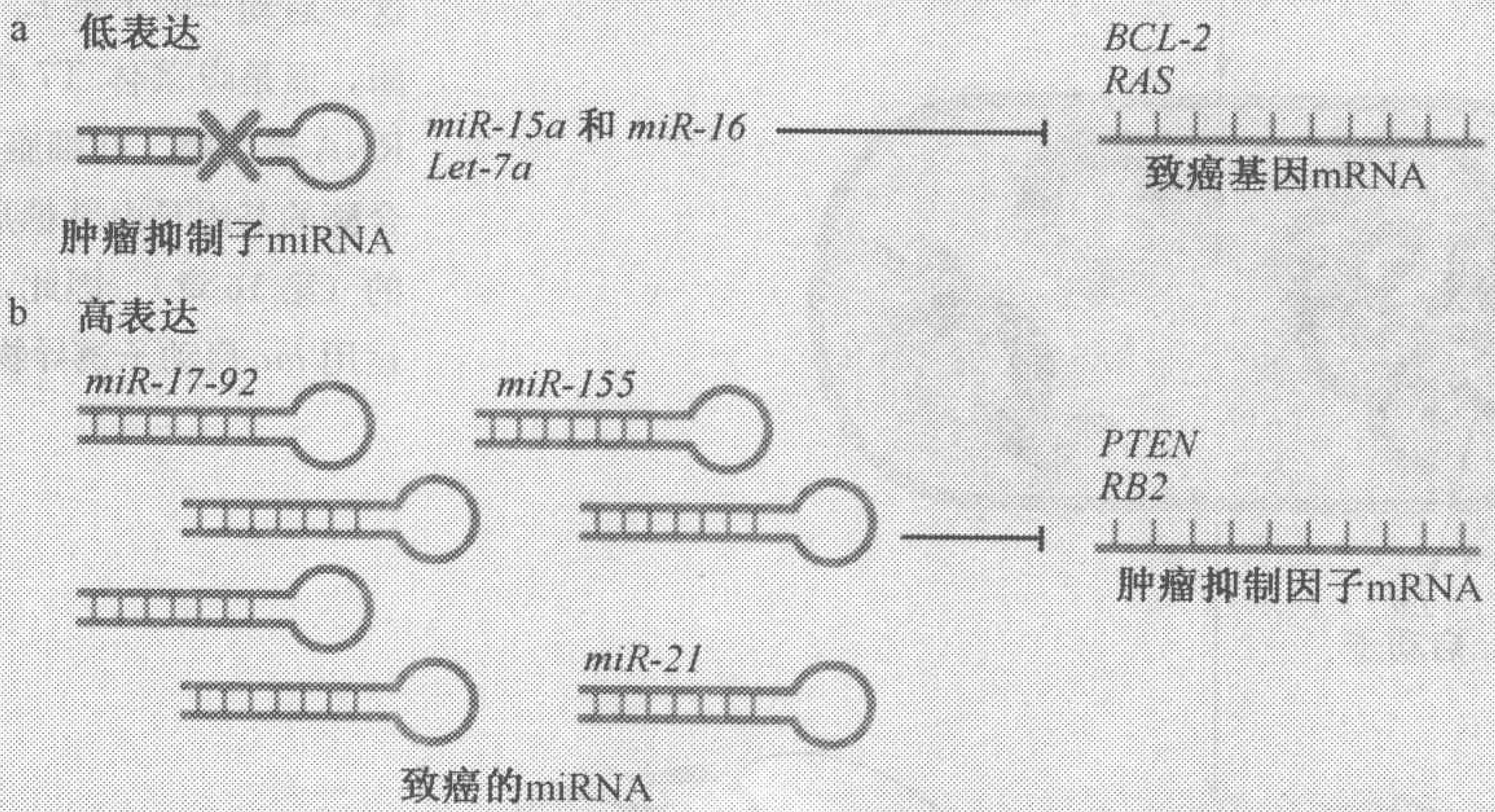
在人体被确认的 300 个 miRNA 中，一半以上的 miRNA 所处的基因组区域在癌症发生时常被破坏。因此，在许多癌症中，根据染色体重排性质的不同，这些 miRNA 的基因要么缺失要么扩增。因此，举例来说，miRNA *miR-15* 和 *miR-16* 通过下调 *BCL2* 基因，诱导细胞凋亡（框 18-3 图 1）。在西半球，最常见的成人白血病是慢性淋巴细胞白血病（CLL），这是一种与 13 号染色体上的一个区域（13q14）缺失相关的疾病。miRNA *miR-15* 和 *miR-16* 基因位于基因组的这个区域，也只有这两个基因确实位于与 CLL 相关的最小缺失内。因此，当这些基因缺失后，凋亡被下调，肿瘤可以更易发生并发展。

在 13 号染色体的另一个区域（13q31）发现了一个致癌的 miRNA：*miR-17-92*。与正常组织相比，这个基因的表达在多种癌症（包括肺癌）中显著升高，特别是在最具侵袭性的癌症中（如小细胞肺癌）。此外，在转基因小鼠中这一 miRNA 的过量表达也会导致肿瘤发生。在 *miR-17-92* 的许多预测的靶基因中，有两个肿瘤抑制基因，*PTEN* 和 *RB2*。一个肯定的靶基因是细胞周期活动因子 E2f1。框 18-3 图 1 示这两个和其他一些癌症中 miRNA 的实例。

脆性 X 智力低下

RISC 复合物的生物化学分析鉴定了很多相关蛋白质。其中一种是脆性 X 智力低下蛋白（fragile X mental retardation protein，FMRP）。这种蛋白质的编码基因与 X 染色体连锁，它的突变是造成最常见的可遗传的智力低下的原因。FMRP 是一个和基因调控有关的 RNA 结合蛋白。因基因表达受干扰，FMRP 缺陷的患者智力低下，还具有很多其他的发育缺陷。

果蝇有一个 FMRP 的同源基因。在这种基因缺陷的果蝇中，常会观察到连接神经元和肌肉间的突触异常。果蝇的一种 Argonaute 蛋白被发现与 FMRP 协同作用，另外一些有关 Argonaute 的研究表明 FMRP 与 RNAi 机器中的这一组分结合。在人类细胞中也有相似的发现，显示这种脆性 X 与 RNAi 基因调控通路之间存在吸引人的联系。



框 18-3 图 1 miRNA 作为致癌基因的抑制因子。a. 据这一模型，一个在正常情况下下调致癌基因的 miRNA 具有肿瘤抑制基因的功能。例如，通过突变或缺失而使 miRNA 失去功能就有可能导致靶致癌基因的异常表达，从而使肿瘤形成。b. 在这里，miRNA 的扩增或过量表达可以通过刺激增殖、促进血管生成和侵袭而使肿瘤形成，这种 miRNA 可以下调肿瘤抑制因子或其他在分化中起作用的重要基因。（Garzon R. et al. 2006. Trends. Mol. Med. 12：580-587，Fig. 2. ©Elsevier. 经许可重绘）。

RNAi 已成为基因表达操作的有力工具

RNAi 是研究者在试图操作基因表达的过程中观察发现的（框 18-2）。在植物的协同抑制（cosupression）和线虫的反义 RNA 的研究中，对出现的未知现象的探索导致了 RNAi 的发现。因此，一点也不奇怪，一旦被理解，RNAi 就迅速被开发成基因表达操作的工具。在线虫中 RNAi 的使用已成常规。已有编码 dsRNA 的文库，可用来研究线虫基因组中的任一靶基因。这些文库可以被用来筛选任一特定基因的表达被抑制后的线虫。图 18-14 显示了这一过程。线虫以细菌为食，在实验室中，用 *E. coli* 喂养线虫。实验显示，最快捷的整合进线虫基因组的途径就是它的胃：任何所需要的 dsRNA 都可以在喂养线虫的 *E. coli* 中表达，并可为线虫细胞中待启动的 RNAi 提供足够的底物，关闭与原有 dsRNA 同源的基因。

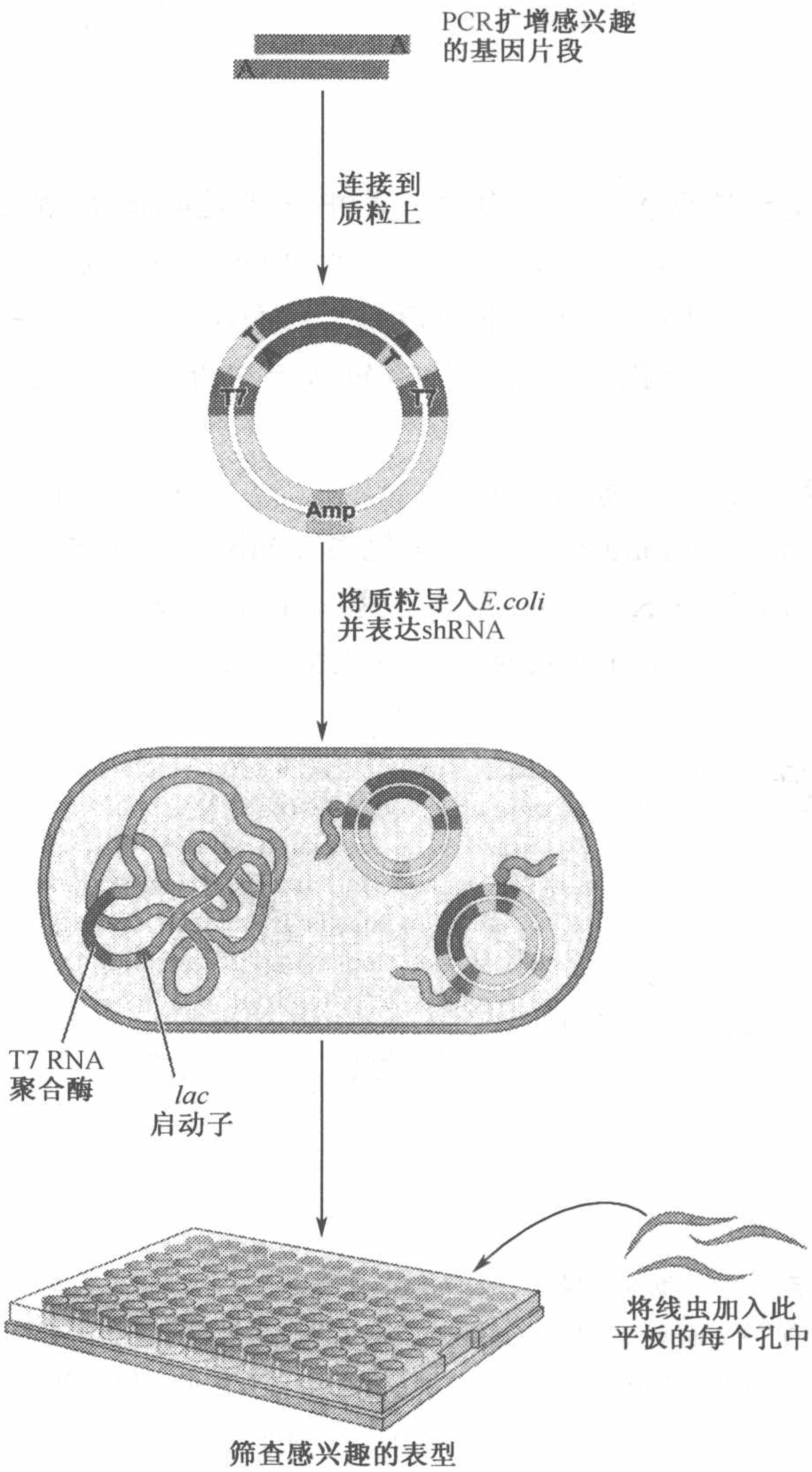


图 18-14 线虫的 RNA 干扰（RNA interference）可通过喂养表达 dsRNA 的细菌来诱导。本图开始的几个步骤所需要的分子操作的细节请参见第 21 章。质粒的 shRNA 表达受启动子的控制，识别这一启动子的不是 *E. coli* 的 RNA 聚合酶，而是噬菌体 T7 中的一个单亚基的 RNA 聚合酶。在细胞中人工表达这个聚合酶的基因是由乳糖启动子（*lac*）控制的（第 16 章）。因此，shRNA 的形成可运用 *lac* 启动子诱导物进行控制。

显而易见，这种方法对在哺乳动物细胞中进行基因筛选非常有帮助，而这是传统的遗传筛选做不到的。现在 siRNA 已经能够在试管中人工合成：并能导入哺乳动物的培养细胞中，引发 RNAi 反应，使靶基因表达下调，但转染（将 RNA 转入细胞）的效率很低。较长的 dsRNA 分子使用起来也有问题，因为它们引发的反应会关闭细胞中所有的翻译。这个反应机制起源于封闭病毒复制的进化过程，而很多病毒有 RNA 基因组。尽管存在这些缺点，这一方法还是可以被有效地运用。

让细胞表达（而不是高浓度地导入）dsRNA，不仅可以解决上述问题，还有另一个好处：这些 RNA 可以持续生成，而非如转染后短时间表达。但这种方法也有其难以克服的困难：如何保证两条互补链在同一个细胞中都被表达，并找到对方形成 RNAi 的有活性的 dsRNA 底物？

研究者发现模拟 miRNA，要比将 dsRNA 转入细胞更有成效。为此，短基因以合成的寡核苷酸克隆进质粒并建立了文库。每一个短基因被设计成可折叠成柄-环结构的转录物。它们被 dicer 加工形成 siRNA，直接沉默靶基因。这种合成的短基因被称为“短发卡 RNA”（short hairpin RNA，shRNA）基因。设计恰当的 shRNA，任何一个靶基因都可以研究；使用合适的文库，任何一个遗传筛选都可以进行。例如，在这样的文库中，每个质粒都编码一个作用于不同基因的 shRNA。然后整个文库都转染进细胞中，每个细胞都接受了一个不同 shRNA。通过筛选特定表型的细胞，就可以鉴定那个被抑制即导致该表型的靶基因。

调控 RNA 与 X 染色体灭活

X 染色体灭活产生镶嵌个体

雌性哺乳动物有两条 X 染色体，而雄性则有一条 X 和一条 Y 染色体。虽然这是性别决定的基础并使雌雄有别，但这依然带来了麻烦：假如不加检测，雌性 X 染色体编码基因的水平将是雄性的两倍。这一不平衡的表达将有可能破坏代谢和细胞过程。为了避免这一问题，需要一种剂量补偿（dosage compensation）的机制。在哺乳动物中，雌性通过灭活其中一条染色体基因的表达来达到这一目的，那条染色体上的基因因而差不多没有一个能够表达。在胎生哺乳动物发育过程中，灭活发生于 32 到 64 细胞期。这个时候，在每一细胞中，究竟哪条 X 染色体灭活，母源的或是父源的，其选择显然是随机的。一旦细胞选择了灭活的 X 染色体，这一细胞产生出来的所有后代都将维持同一条 X 染色体的灭活。

各个细胞中 X 染色体灭活的随机性导致的后果是雌性都是镶嵌体（mosaic）。一些细胞中表达父本的 X 染色体，而另外一些细胞则表达母本的 X 染色体。通常来说，这是无关要紧的，尽管这可能影响 X 连锁疾病综合征的严重程度。这一程度依赖于表达或沉默突变基因的细胞的比例。一个广为人知的例子是 calico 猫（“玳瑁猫”）（图 18-15）。猫的 X 染色体上的一个基因影响毛色是橙色的或是黑色的。一个等位基因决定

毛色为橙色，而另外一个等位基因决定毛色为黑色。杂合子的猫显示不同的橙黑斑块毛色，说明该处的细胞那一条 X 染色体被灭活。这一现象也解释了为什么所有的 calico 猫都是雌性的。白猫则决定于一个常染色体的基因。

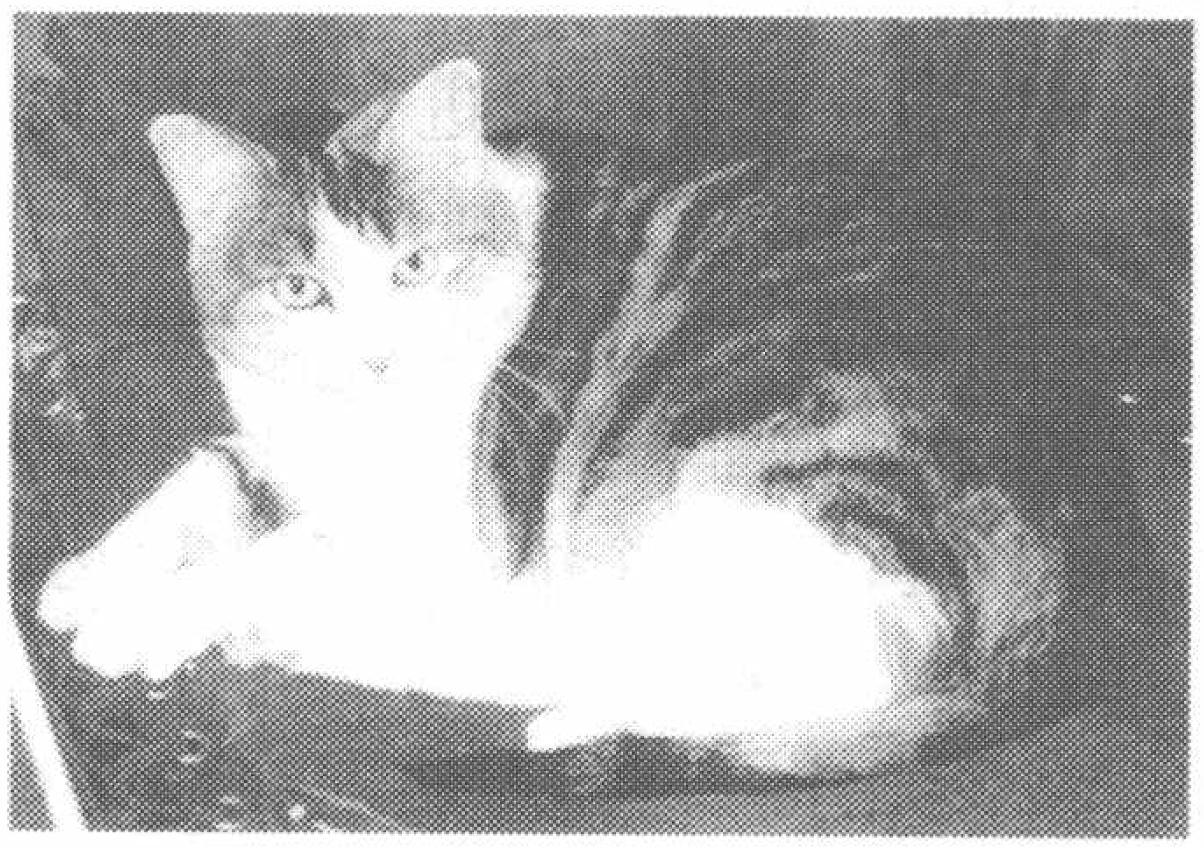


图 18-15 X 染色体灭活的实例观察：calico 猫。如文中所述，橙黑斑块的毛色间接显示了 X 染色体的灭活。（VG 惠赠.）

Xist 是雌性哺乳动物中灭活一条 X 染色体灭活的调控 RNA

X 染色体是如何灭活的？灭活又是如何在发育过程中遗传下去？启动灭活的 RNA 分子称为 Xist。Xist RNA 是由一个对 X 染色体灭活至关重要的位点——X 染色体灭活中心（X-chromosome inactivation center，Xic）编码的。*Xist* RNA 包被在表达该基因的 X 染色体，如图 18-16a 中的原位杂交所示。到底是什么原因致使包被发生，为什么包被仅限于一条 X 染色体，即为什么仅以顺式（*cis*）起作用？然而，已知 Xist 的作用是灭活的关键，而且不需要 Xic 之外的别的 X 染色体序列。当 Xist 异位表达于一个常染色体位点的时候，如一个非性染色体，Xist 能够在不同的程度上沉默这条染色体上的基因，也就是“灭活”表达它的那条常染色体。

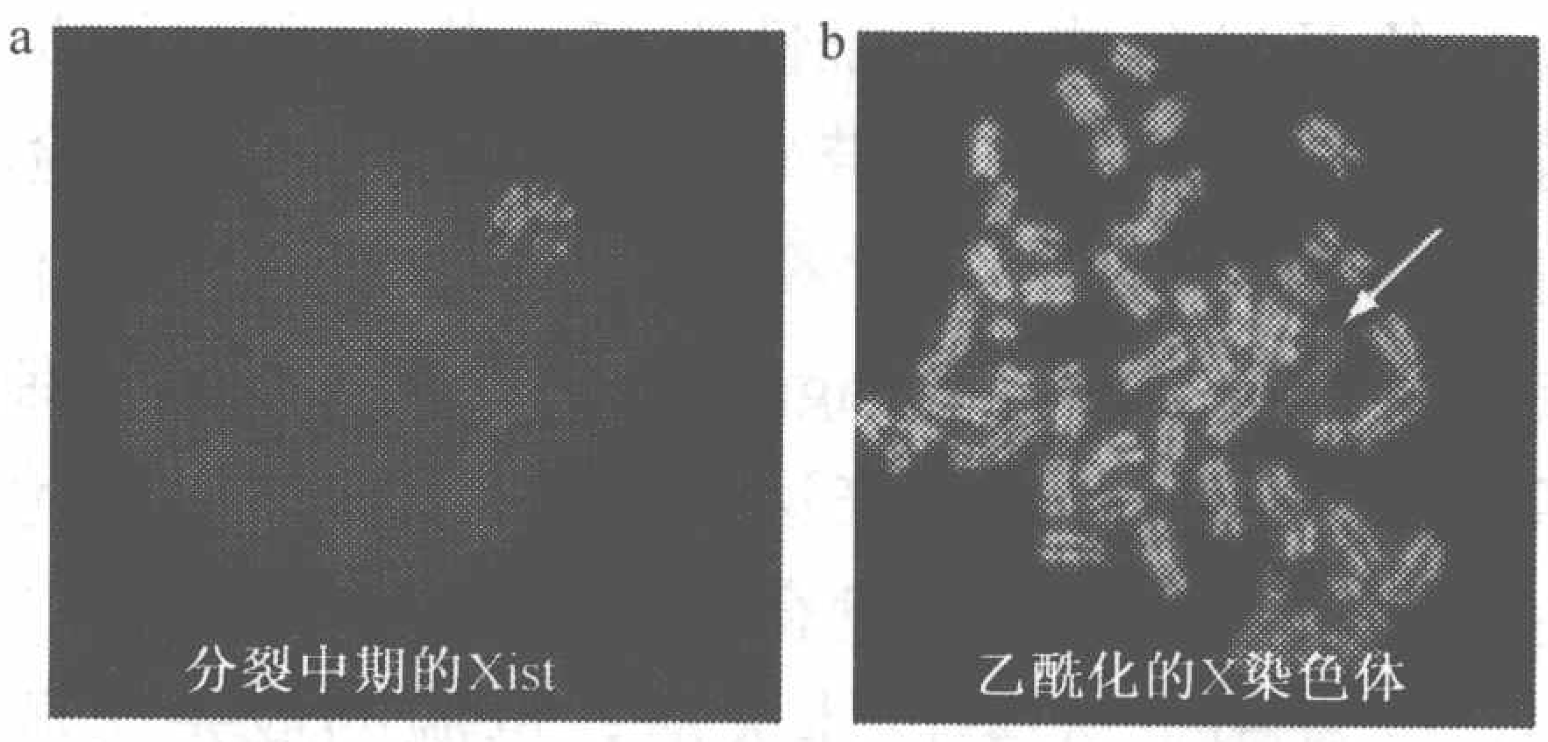


图 18-16 X 染色体灭活的实例观察：分子标记。a. 中期细胞的原位杂交显示灭活的 X 染色体上 *Xist* RNA 的定位。b. 染色体染色以显示组蛋白 H4 乙酰化的。箭头所指为灭活的 X 染色体，其乙酰化修饰水平要比其他染色体低得多。（经许可复制自 Brockdorff N and Turner B. M. 2007. Epigenetics [ed. Allis et al.]，p. 327. © Cold Spring Harbor Laboratory Press.）

Xist RNA 本身并不能引发沉默，但它聚集了其他修饰和浓缩染色质的因子，这些因子也可能使 DNA 甲基化（第 17 章中其他基因沉默的例子）。正是这些修饰引发了沉默并保证了沉默的遗传。这些修饰一旦确立，*Xist* 本身就不再需要了。图 18-16b 显示灭活的 X 染色体组不同的蛋白质修饰。在这里，与基因组其他部分相比，这条灭活的 X 染色体的乙酰化程度要低很多。我们在前几章（第 7 章和第 17 章）中所看到的，去乙酰化的组蛋白与基因组中不转录的区域相关。

细胞是如何选择将哪个 X 染色体灭活呢？这一答案现在还不清楚。但另一个调控 RNA 可能是关键。这个 RNA 也是由 *Xic* 位点编码的，但它在另一条 DNA 链上并与 *Xist* 基因重叠。它被称为 *Tsix*（将 *Xist* 反过来拼写），起负调控 *Xist* 的作用。事实上，如果某一条 X 染色体上的 *Tsix* 发生了突变，这条染色体将被选为灭活的染色体。因此，*Xist* 和 *Tsix* RNA 生成和稳定的平衡将在每一个细胞中导致这样或那样的结果。

与哺乳动物一样，剂量补偿在所有动物中都是必需的，包括线虫和果蝇的例子。但在不同的情况下，补偿的机制是不同的。例如，在果蝇中，剂量补偿是通过增强雄性的 X 染色体连锁定基因的表达（而不是降低雌性的表达）来实现的。但是在这里，非编码的调控 RNA 也参与了这一机制。在这个过程中，RNA（称为 *roX1* 和 *roX2*）参与将染色质修饰结合体聚集到雄性 X 染色体上的基因而帮助激活转录。

在小鼠的某些印迹基因簇中也发现了非编码 RNA。我们已在第 17 章中讨论过基因表达的印迹。在已讨论过的机制的基础上，现在相信认为在一些情况下调控 RNA 也部分参与了这一过程，它们所起的作用可能类似于 X-灭活中的 *Xist* 和 *Tisx*。这种非编码 RNA 是由在各个基因簇中发现的印迹控制区域（imprinting control region，ICR，图 17-26）调控的，它们沉默同一染色体上的邻近基因。DNA 和组蛋白的甲基化都参与这一沉默机制。

小结

尽管早在 1961 年便已提出 RNA 分子有可能是基因调控因子，但直到近几年，RNA 作为调控因子的普遍存在和重要作用才被阐明。在这之前，只有 *E. coli* 的色氨酸操纵子（*trp* operon）的衰减作用这一个罕见的例子。在那个例子中，已知 mRNA 的 5′区 RNA 序列调控下游基因的表达，这个 RNA 区域中另一种方式的分子内碱基配对形成的另一种不同的二级结构，对基因产生不同的影响。一种构象中，在进入下游基因的编码区之前转录就被终止，而在另一种构象中，转录能够继续，基因能够表达。这些基因编码色氨酸合成相关的酶，其是否表达取决于待用的色氨酸空载 tRNA：如果细胞中空载 tRNA 很少，那么色氨酸就被需要而基因就需表达；如果细胞中含有大量的负载色氨酸 tRNA，*trp* 基因的转录就被终止。

核糖开关调控基因的机制与此相似：基因的 5′-非翻译区二级结构的不同决定这些基因是否转录（或者在一些情况下，决定翻译是否启动）。在核糖开关的作用下，选择哪一种二级结构取决于它与控制那个基因的 RNA 配体的直接结合。它们都是小的代谢

产物（维生素 B_{12}、SAM、氨基酸）和基因，这些基因分别受控，编码那些使用这些代谢产物的酶。

E. coli 也编码以反式调控基因的小 RNA（称为 sRNA）。这些 sRNA 是由小基因编码的，与 mRNA 的互补序列形成碱基配对。这种情形要么抑制这些靶 mRNA 的翻译，导致它们的降解，要么在某些情况下促进它们的翻译。细菌的小 RNA 的作用在很多方面与真核细胞中的调控基因的小 RNA 相似，尽管真核生物生成调控 RNA 的机制和它们对靶基因产生影响的机制与细菌的很不一样。

真核生物将 RNAi 中的 dsRNA 作为基因沉默剂。特殊的酶（Drosha，更多情况下是 Dicer）识别 dsRNA，并以短 RNA（长 21～22 核苷酸）生成基因沉默的活性因子。dsRNA 底物是由两条互补链的碱基配对形成的，或者通过单个分子的折叠形成具有特定双链区的二级结构。而前一种情况是由 Dicer 作用的，生成的短产物称为 siRNA；后一种需要 Dicer，然后是由 Drosha 作用的，产物被称为 miRNA。Dicei 和 Drosha 都含有核酸酶的结构域，RNA 底物的剪切是以 RNA 长度和结构，而非特殊的序列为基础的。

siRNA 和 mRNA 一旦形成，其功能在本质上是相同的。它们被整合进称为 RISC 的机器中，其中一条 RNA 链被选作“指导 RNA”，指导成熟的 RISC 与靶 RNA 和互补的指导 RNA 结合。一旦结合，RISC 要么沉默 RNA（通过催化含有 RNAase 相关结构域的底物 Argonaute），要么抑制 mRNA 的翻译。选择哪种方式的沉默基本上取决于指导 RNA 和靶 RNA 的碱基配对的程度——配对越好，越有可能引发沉默。指导 RNA 也可以指导 RISC 的相关组蛋白修饰复合体与启动子区相结合，通过修饰启动子而在转录水平上沉默基因。即使在这些情况下，与启动子相结合也很有可能是通过指导 RNA 与 mRNA 的碱基配对达到的，但在这个例子中，RNA 仍然形成，因此与这个基因上的 RNA 聚合酶Ⅱ有关。

miRNA 是由生物体内的基因编码的，在发育过程中它们起基因调控的作用，这在线虫和植物中已有研究得很好的例子。miRNA 也与癌症相关，有些 miRNA 被归类为抑癌基因，有的被归类为致癌基因。形成 siRNA 的 dsRNA 来源不一，它们既可能来自侵染的病毒，也可能来自重复区域的转录（着丝粒或转座子），或是有意导入到细胞中的（在希望下调特定基因表达的实验中）。RNAi 的使用已成为常规工具，在传统遗传学不起作用的系统中特别有意义。

最后，我们讨论了调控重要生物学过程的非编码 RNA 的一个实例。雌性动物有两条 X 染色体，而雄性动物只有一条 X 染色体（和一条 Y 染色体）。为了保证两种性别产生大致相同数量的 X 染色体基因产物，这种不等的染色体数量必须由剂量补偿的机制来纠正。哺乳动物通过在雌性中灭活其中一条 X 染色体而实现剂量补偿。X 染色体编码的一种 RNA 分子（*Xist*）调控这一过程。

参考文献

书

Hannon G.J. 2003. *RNAi: A guide to gene silencing.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Regulatory RNAs. 2006. *Cold Spring Harbor Symp. Quant. Biol.*, vol. 71. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

细菌小RNA

Coppins R.L., Hall K.B., and Groisman E.A. 2007. The intricate world of riboswitches. *Curr. Opin. Microbiol.* **10:** 176–181.

Gottesman S. 2005. Micros for microbes: Non-coding regulatory RNAs in bacteria. *Trends Genet.* 21: 399–404. Review.

Storz G. and Haas D. 2007. A guide to small RNAs in microorganisms. *Curr. Opin. Microbiol.* **10:** 93–95.

核糖开关和衰减作用

Gilbert S.D. and Batey R.T. 2006. Riboswitches: Fold and function. *Chem. Biol.* **13:** 805–807.

Gollnick P., Babitzke P., Antson A., and Yanofsky C. 2005. Complexity in regulation of tryptophan biosynthesis in *Bacillus subtilis. Annu. Rev. Genet.* **39:** 47–68.

Winkler W.C. 2005. Riboswitches and the role of noncoding RNAs in bacterial metabolic control. *Curr. Opin. Chem. Biol.* **9:** 594–602.

Winkler W.C. and Breaker R.R. 2005. Regulation of bacterial gene expression by riboswitches. *Annu. Rev. Microbiol.* **59:** 487–517.

Yanofsy C. 2000. Transcription attenuation: Once viewed as a novel regulatory strategy. *J. Bacteriol.* **182:** 1–8.

RNAi的机制

Baulcombe D. 2005. RNA silencing. *Trends Biochem. Sci.* **30:** 290–293.

Cook A. and Conti E. 2006. Dicer measures up. *Nat. Struct. Mol. Biol.* **13:** 190–192.

Hannon G.J. and Rossi J.J. 2004. Unlocking the potential of the human genome with RNA interference. *Nature* **431:** 371–378.

Pei Y. and Tuschl T. 2006. On the art of identifying effective and specific siRNAs. *Nat. Methods* **3:** 670–676.

Peters L. and Meister G. 2007. Argonaute proteins: Mediators of RNA silencing. *Mol. Cell* **26:** 611–623.

Rana T.M. 2007. Illuminating the silence: Understanding the structure and function of small RNAs. *Nat. Rev. Mol. Cell Biol.* **8:** 23–36.

Tang G. 2005. siRNA and miRNA: An insight into RISCs. *Trends Biochem. Sci.* **30:** 106–114.

Tolia N.H. and Joshua-Tor L. 2007. Slicer and the Argonautes. *Nat. Chem. Biol.* **3:** 36–43.

Valencia-Sanchez M.A., Liu J., Hannon G.J., and Parker R. 2006. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev.* **20:** 515–524.

Zaratiegui M., Irvine D.V., and Martienssen R.A. 2007. Noncoding RNAs and gene silencing. *Cell* **128:** 763–776.

MicroRNA

Ambros V. 2004. The functions of animal microRNAs. *Nature* **431:** 350–355.

Bushati N. and Cohen S.M. 2007. microRNA functions. *Annu. Rev. Cell Dev. Biol.* **23:** 175–205.

Hannon G.J., Rivas F.V., Murchison E.P., and Steitz J.A. 2006. The expanding universe of noncoding RNAs. *Cold Spring Harbor Symp. Quant. Biol.* **71:** 551–564.

miRNA和RNAi的研究史

Baulcombe D. 1994. Replicase-mediated resistance: A novel type of virus resistance in transgenic plants? *Trends Microbiol.* **2:** 60–63.

Lee R., Feinbaum R., and Ambros V. 2004. A short history of a short RNA. *Cell* (suppl. 2) **116:** S89–S92, 1 p. following S96.

Ruvkun G., Wightman B., and Ha I. 2004. The 20 years it took to recognize the importance of tiny RNAs. *Cell* **116**(suppl. 2): S93-S96, 2 p. following S96.

MicroRNAs和疾病

Calin G.A. and Croce C.M. 2006. MicroRNA signatures in human cancers. *Nat. Rev. Cancer* **6:** 857–866.

Dalmay T. and Edwards D.R. 2006. MicroRNAs and the hallmarks of cancer. *Oncogene* **25:** 6170–6175.

Garzon R., Fabbri M., Cimmino A., Calin G.A., and Croce C.M. 2006. MicroRNA expression and function in cancer. *Trends Mol. Med.* **12:** 580-587.

Zhang B., Pan X., Cobb G.P., and Anderson T.A. 2007. microRNAs as oncogenes and tumor suppressors. *Dev. Biol.* **302:** 1–12.

Xist RNA和X染色体无活

Alexander M.K. and Panning B. 2005. Counting chromosomes: Not as easy as 1, 2, 3. *Curr. Biol.* **15:** R834–R836.

Deng X and Meller V.K. 2006. Non-coding RNA in fly dosage compensation. *Trends Biochem. Sci.* **31:** 526–532.

Heard E. 2005. Delving into the diversity of facultative heterochromatin: The epigenetics of the inactive X chromosome. *Curr. Opin. Genet. Dev.* **15:** 482–489.

Ng K., Pullirsch D., Leeb M., and Wutz A. 2006. *Xist* and the order of silencing. *EMBO Rep.* **8:** 34–39.

Pauler F.M., Koerner M.V., and Barlow D.P. 2007. Silencing by imprinted noncoding RNAs: Is transcription the answer? *Trends Genet.* **23:** 284–292.

第19章 发育和进化的基因调控

本章概要

•

三种不同的策略指导细胞在发育过程中表达不同的基因组合

•

三种确定差异基因表达策略的例证

•

果蝇胚胎发育的分子生物学

•

同源异型基因：一类重要的发育调节因子

人体拥有 200 多种不同的细胞类型，它们都来自于单一细胞——受精卵。这些相同遗传背景的细胞通过在发育过程中不同系列的基因的表达而彼此不同。例如，发育中的肌肉细胞表达特殊形式的肌动蛋白、肌浆球蛋白、原肌球蛋白，而这些蛋白不会出现在其他组织如肝脏和肾脏中。从以下的叙述可见这种差异基因表达的程度。典型的无脊椎动物，如果蝇和线虫拥有大约 15 000～20 000个基因，而脊椎动物则拥有此数的 2 倍多，即 25 000～30 000个基因。全基因组微阵列可检测哪些基因在特定组织的表达。例如，大约 7%～8%（约 1500 个基因）的秀丽新小杆线虫的基因会在肌肉组织中表达（图 19-1）。不同类型的细胞，如肌肉细胞和神经元细胞会表达不同但又部分重叠的一系列基因。通常一种细胞表达的基因有不到一半可见于另一种细胞，而决定每种细胞特异性的标志基因也不过100 个（框 19-1）。

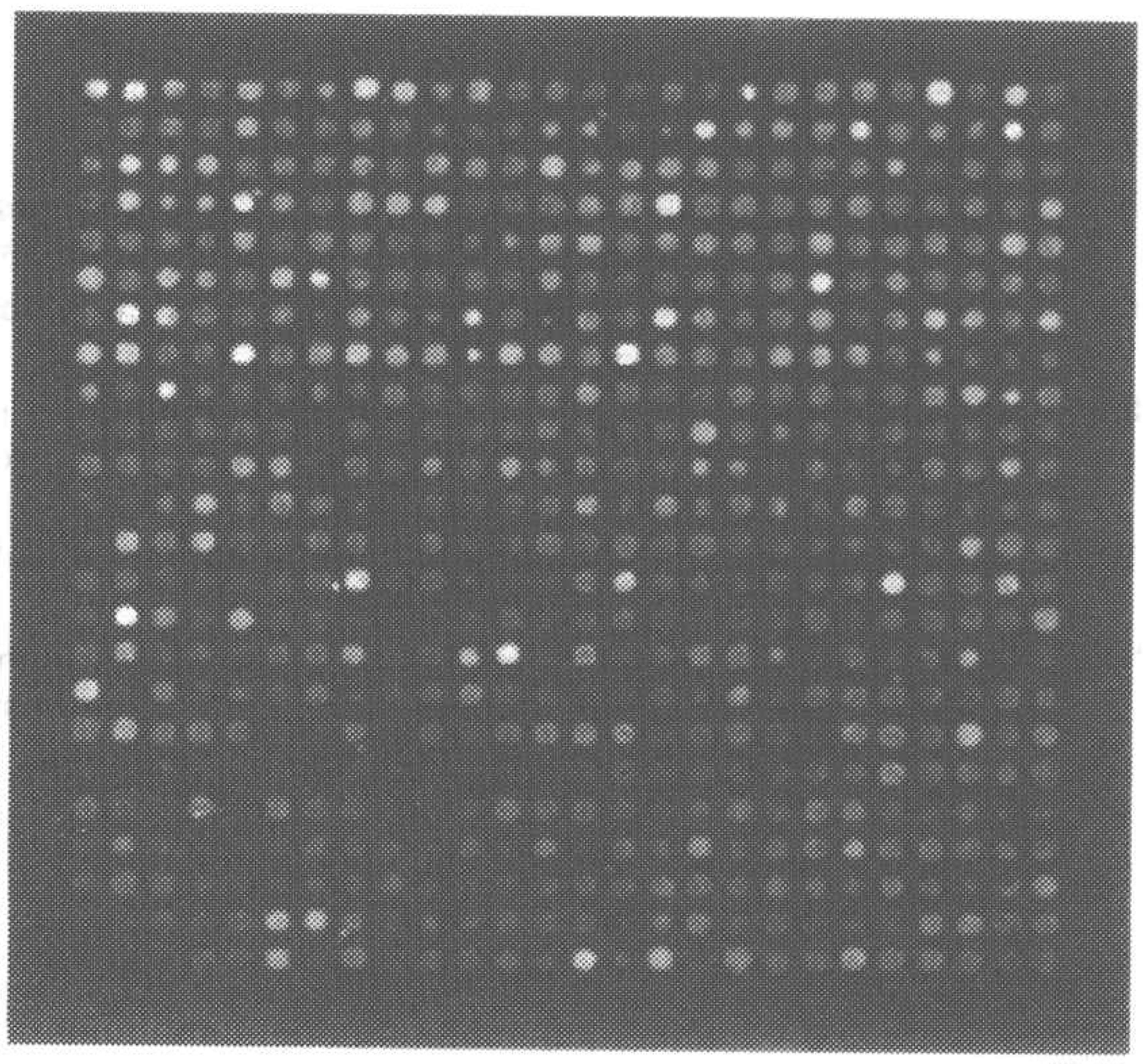


图 19-1 微阵列比较线虫两种组织（肌肉和神经元）的基因表达模式。 每个格点包含一个线虫基因的短 DNA 片段。RNA 从肌肉和神经元组织提取后分别以红、绿两色荧光染料标记。因而，红点代表表达肌肉的基因，绿点代表表达于神经元的基因。黄点则表明基因在两种组织都有表达。由此看出两种组织表达了不同的基因组合。（来源：Courtesy of Stuart Kim.）

框 19-1 微阵列：理论和操作

微阵列提供了基因组范围基因表达的概况。微阵列通常包含在显微玻片上固化的已知核苷酸序列，可以进行一系列并行的杂交实验。要制备微阵列，可先经聚合酶链反应（PCR，第 20 章）扩增蛋白质编码序列。最常见的扩增方式涉及运用短的寡聚核苷酸序列（常规约为 20bp）划定基因组中一个基因的一个外显子。然后，成对的寡聚核苷酸，每对包含一个外显子，与基因组 DNA 杂交并通过 PCR 扩增。扩增后的基因组 DNA 以一系列的斑点贴附在玻片上。玻片上的斑点代表一系列特定的蛋白质编码序列的扩增片段。显微玻片可以包含 40 000 个这样的 PCR 片段，这样一张玻片就可以装载人类基因组的全部蛋白质编码基因。

要研究全基因组范围的基因表达时，可将玻片与不同荧光标记的 RNA 探针杂交。图 19-1 比较了线虫肌肉和神经元的基因活性。组织总 RNA 经提取后由不同的染料标记，如肌肉 mRNA 标记为红色，神经元 mRNA 标记为绿色。然后这些标记的 mRNA 在同一张玻片上同时与线虫的约 2 万个基因的 PCR 扩增片段进行杂交。如果两种 mRNA与同一斑点杂交，该位点将会呈现黄色，表明该位点代表的基因在两种组织同时表达。斑点红色更强表明肌肉组织的基因表达更多。反之，则表明基因在神经元表达。

这一方法可以用于比较任何两种组织的基因表达概况。大量的研究集中于比较正常和肿瘤组织的基因表达，还可以从正常的酵母或果蝇分离 RNA 并与突变的个体进行比较。

那么来自于同一受精卵的细胞是如何建立不同的基因表达程序的呢？大部分差异基因表达是在转录水平调节的，前两章描述了这种表达的基本机制。本章前半部分我们将描述细胞在发育过程中是如何通讯以确保正常发育所需的一套特定基因的表达。这些调节机制都会给出例子。后半部分我们将描述这些机制是如何与第 17 章谈到的转录调节机制结合来调节整个机体，如果蝇的发育。在本章的最后，我们将讨论基因调控的改变是如何造成进化中的动物形态各异的，另外还将描述一类极为重要的发育控制基因——Hox 基因。

三种不同的策略指导细胞在发育过程中表达不同的基因组合

我们已经看到基因表达是如何受到细胞从环境接收的信号的调节。例如，乳糖在大肠杆菌中活化乳糖操纵子，在哺乳动物病毒感染激活 β-干扰素的表达。本章将集中描述遗传背景相同的细胞表达不同的基因进而分化为不同的细胞类型的策略。三种主要策略是：mRNA 定位（mRNA localization）、细胞-细胞接触（cell-to-cell contact）及分泌信号分子扩散传导（signaling through the diffusion of a secreted signaling molecule）（图 19-2）。每种策略简述如下。

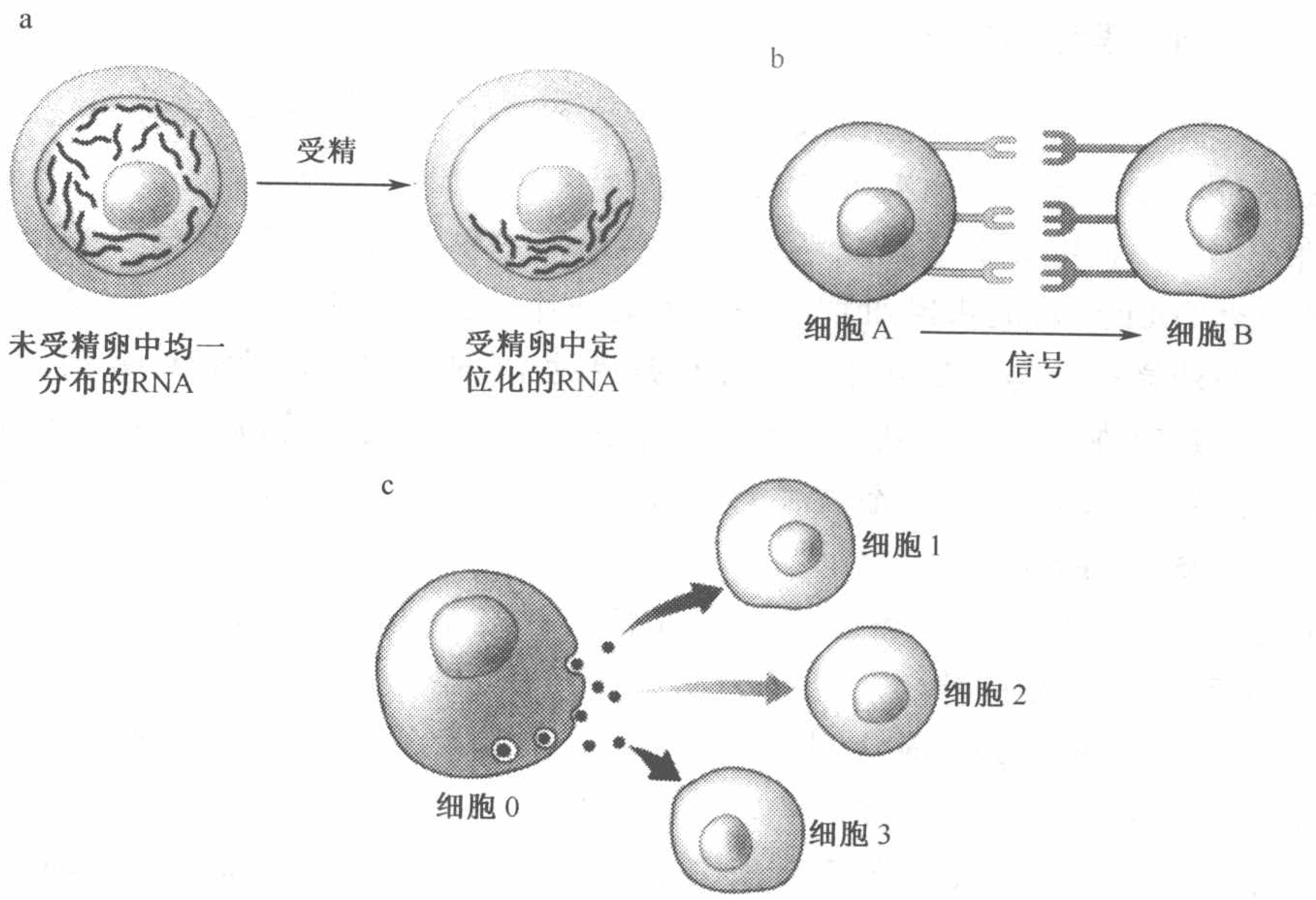


图 19-2 发育过程中启动差异基因活性的三种策略。a. 在某些动物，卵细胞的部分母系 RNA 在受精前后定位化。在这个例子中，特定的 RNA（绿花线）定位于卵细胞的植物极（底部）。b. 细胞 A 必须与细胞 B 直接反应激化细胞 B 表面的受体。这是因为细胞 A 的配体“系”在细胞 B 的细胞膜上。c. 在这个例子中，细胞表现出远距离信号传导。细胞 0 分泌信号分子并扩散至细胞间质。不同的细胞（1、2、3）接收这些信号并最终产生基因活性的改变。

由于细胞骨架固有的极性使某些 RNA 在卵细胞和胚胎中被定位化

一种建立两个相同遗传背景的细胞不同特性的策略是在细胞分裂过程中使关键的调节分子非对称性分配，即子细胞得到不同量的调节分子从而在发育过程中经历不同的过程，通常这种非对称分布的分子是 mRNA。这些 mRNA 为 RNA 结合蛋白和信号传导分子编码，更为常见的是编码转录活化子和抑制子。尽管这些蛋白质产物的功能多种多样，但 mRNA 的定位具有共同机制。通常，它们沿着细胞骨架成分肌动蛋白或微管蛋白运输。mRNA 分布的非对称性来自于这些成分的非对称性。

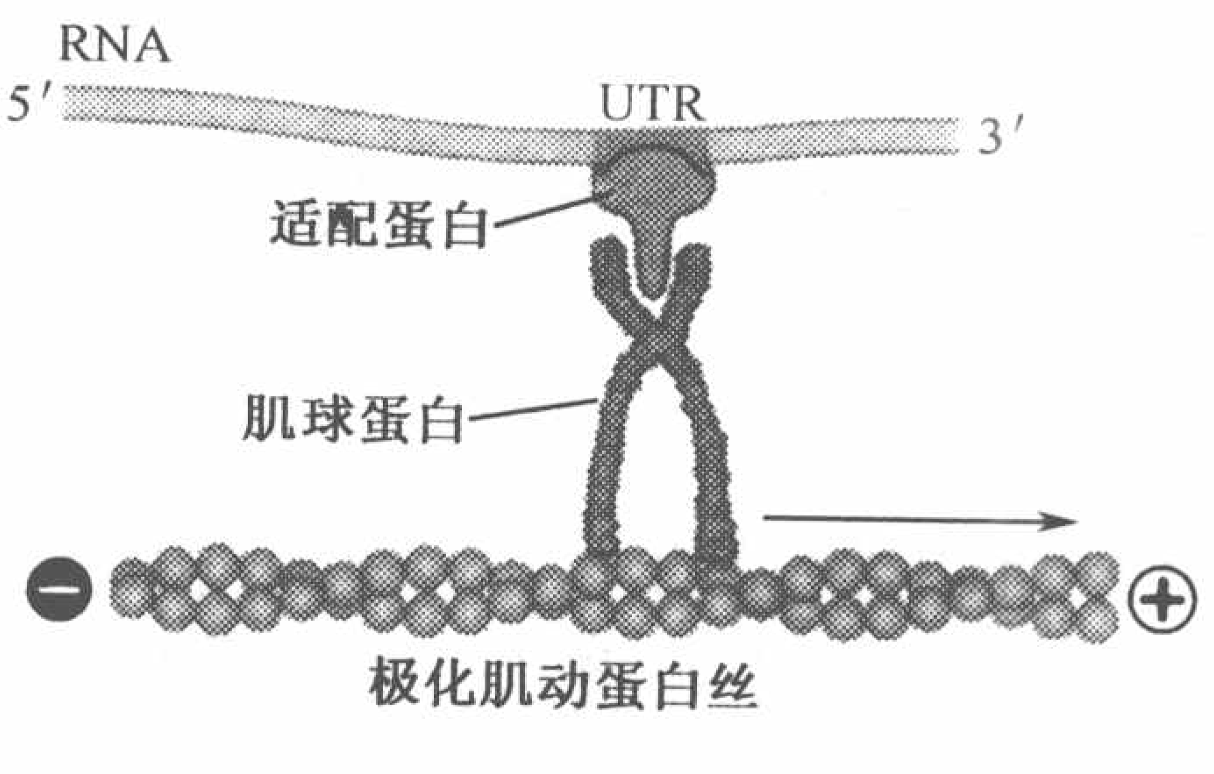


图 19-3 适配蛋白与 mRNA 3′-端非翻译区（3′-UTR）的特定序列结合。适配蛋白还与肌球蛋白结合，后者沿着肌动蛋白丝定向由负端向延伸中的正端移动。

肌动蛋白和微管具有内在的极性，表现为正端的定向生长（图 19-3）。mRNA 可经由结合其 3′ 端非翻译区（3′-untranslated trailer，3′-UTR）的特定序列的适配蛋白（adapter protein）从细胞的一端转运到另一端。适配蛋白包含两个结构域，其一结合 mRNA 的 3′-UTR，另一个结合细胞骨架的

特定组分，如肌球蛋白。依据适配蛋白的不同，它们或在肌动蛋白上蠕动或随肌球蛋白生长而移动。以下，我们将会看到这些基本过程是如何使mRNA在卵细胞中定位或者如何使mRNA限定于有丝分裂后的子细胞中。

细胞-细胞接触和细胞分泌的信号分子都会激发相邻细胞基因表达的变化

一个细胞可以通过产生细胞外信号蛋白来影响相邻细胞的基因表达，这些蛋白质合成后或者沉积在细胞膜上，或者分泌到细胞间质中。这两个过程具有共同特点，因此在这里一并叙述。随后，还会讲述到分泌信号的其他作用。

特定信号经由接受细胞表面的受体识别，受体在结合了信号分子后，启动接受细胞基因表达的变化。这种细胞表面受体到细胞核的通信常常涉及信号传导途径（signal transduction pathway）。这一内容在第 17 章已经涉及，其要点概述如下。

有时，配体和受体的结合会触发一系列酶反应并最终修饰核内的调节蛋白（图 19-4a）。或者，活化的受体造成细胞表面或细胞质的 DNA 结合蛋白释放并进入细胞核（图 19-4b）。这些调节蛋白与特定的 DNA 识别序列结合然后活化或者抑制基因表达，配体结合还可以造成受体的蛋白水解剪切。受体剪切后胞质部分从细胞膜释放进入细胞核并与 DNA 结合蛋白结合从而调节这些蛋白质相关的基因（图 19-4c）。例如，这些转运的蛋白质可以从转录抑制子转变为转录活化子，这样，此前被抑制的基因就被激活了，我们将在本章给出相应的例证。

在细胞表面的信号分子只控制那些与其有直接的、物理接触的细胞的基因表达。我们将这类过程归为细胞-细胞接触。反之，分泌于细胞间质的信号分子可以通过长距离发挥作用。有些通过 1～2 个细胞半径的距离，还有一些则可以达到 50 个或更多个细胞的距离。长距离信号分子有时对位置效应有作用，这将在下一节讲到。

分泌信号分子梯度可以引导细胞遵循位置性的不同发育途径

发育生物学的主题之一是位置在细胞发育中对发育中的胚胎或者器官的重要性。定位在果蝇胚胎前部［也就是前区（anterior region）］的细胞将形成成体果蝇的头部，如触角或脑，但不会长成后部（posterior）结构，如腹部或性器。青蛙胚胎的顶部或背部（dorsal）表面会发育成蝌蚪或成年蛙的脊柱但不会变成腹部组织（ventral），如胃。这些例子表明细胞的归宿受到它们在发育胚胎位置的限定，定位在发育中的影响叫做位置信息（positional information）。

建立位置信息的最常见的方式涉及第 17 章所述策略的简单延伸，即对分泌信号分子的运用（图 19-5）的延伸。一小部分细胞的分泌信号分子形成细胞外的梯度（extracellular gradient）（图 19-5a）。接近分泌细胞的那些细胞接受高浓度的分泌蛋白，发育成特定的细胞类型。那些相距较远的细胞则接受较低浓度的信号分子而经历不同的发育途径，控制位置信号的信号分子亦被称为成形素（morphogen）。

定位于接近成形素的细胞接受高浓度的信号分子，因而得到所结合受体的最大活化；而远离信号源的细胞接受低浓度的信号分子，只有部分细胞表面的受体得到活化。假设有 3 排细胞接近成形素，第一排细胞有 1000 个受体得到活化，第二排有 500 个受体，最外一

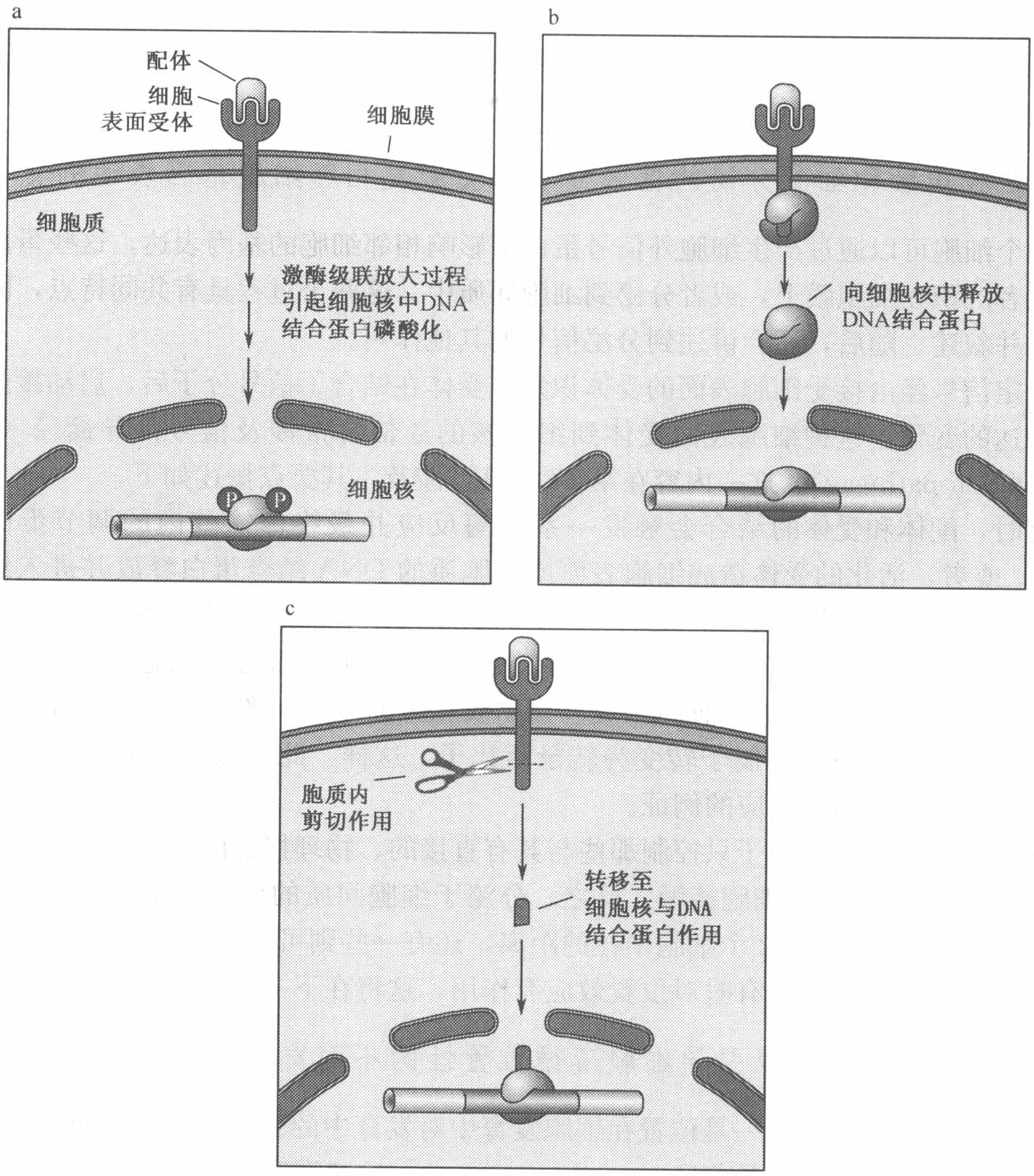


图 19-4 信号传导的不同机制。配体（或称信号分子）与细胞表面受体结合。a. 活化的受体诱发潜伏的细胞激酶并最终造成细胞核内 DNA 结合蛋白的磷酸化。磷酸化造成调节蛋白活化或者抑制特定基因的转录。b. 活化的受体使细胞质中休眠的 DNA 结合蛋白得到释放并进入细胞核，在细胞核内调节蛋白活化或者抑制特定基因的转录。c. 活化的受体受到细胞质中蛋白水解酶的剪切，使受体 C 端部分进入细胞核与特定的调节蛋白反应。产生的蛋白质复合体活化特定基因的转录。

层只有 200 个（图 19-5b）。受体占位的不同程度直接导致接收细胞的差异基因表达。

如上所述，信号分子与细胞表面受体的结合以不同方式引发细胞核内特定的活化转录因子浓度的增加。每种受体控制特定的转录调节因子进而控制特定基因的表达。细胞表面受成形素活化受体的数量决定了转录调节因子在核内出现的数量。最接近成形素的细胞将拥有高浓度的核转录活化因子（图 19-5c）；相反，远离信号源的细胞只有中等或低浓度的转录活化因子。因而，细胞表面活化受体的数量与核内转录活化因子的量成对应关系。那

么，不同浓度的同样的转录调节因子是如何在不同细胞中调节不同的基因表达的呢？

在第 16 章中，我们学到了 λ 抑制子水平的少量变化如何确定感染细菌细胞是溶解或者溶原。同样，成形素浓度的少量变化进而导致核内转录调节因子的变化确定了细胞的性质。具有高浓度已知转录调节因子的细胞可以激活那些在中浓度或者低浓度信号分子细胞激活的基因（图 19-5d）。由调节蛋白的浓度确定基因表达的差异调节是发育生物学中最重要的、贯穿始终的机制。为此，本章将提供多个例证。

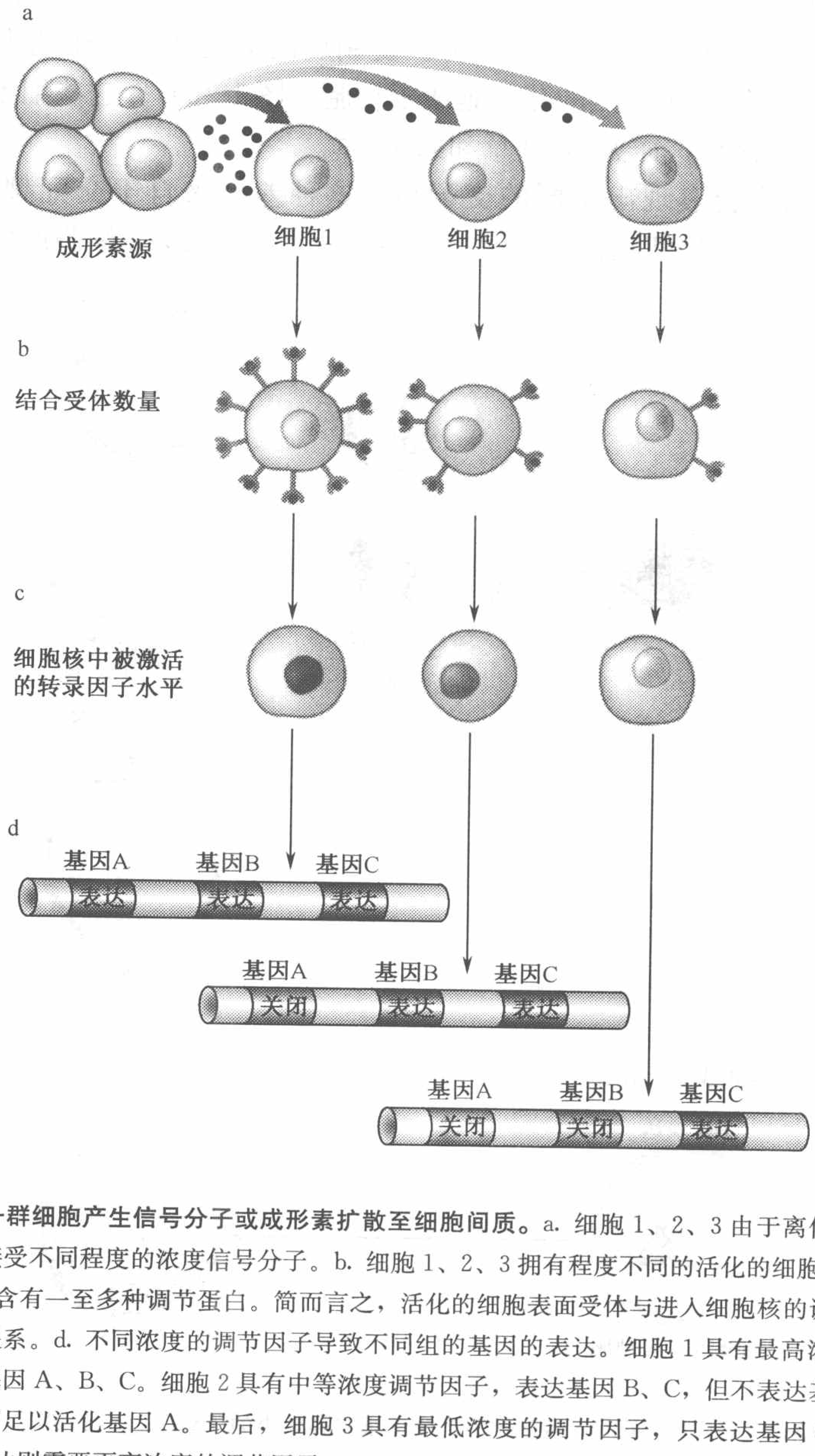


图 19-5　一群细胞产生信号分子或成形素扩散至细胞间质。 a. 细胞 1、2、3 由于离信号源的距离不同而接受不同程度的浓度信号分子。b. 细胞 1、2、3 拥有程度不同的活化的细胞表面受体。c. 3 种细胞含有一至多种调节蛋白。简而言之，活化的细胞表面受体与进入细胞核的调节因子的量呈线性关系。d. 不同浓度的调节因子导致不同组的基因的表达。细胞 1 具有最高浓度调节因子，表达基因 A、B、C。细胞 2 具有中等浓度调节因子，表达基因 B、C，但不表达基因 A，这个浓度还不足以活化基因 A。最后，细胞 3 具有最低浓度的调节因子，只表达基因 C，而基因 A、B 的表达则需要更高浓度的调节因子。

三种确定差异基因表达策略的例证

Ash1 受体定位导致 *HO* 基因关闭并控制酵母交配类型

在叙述动物胚胎 mRNA 定位之前，先来看一看相对简单的酵母的情形。酵母可以以出芽的方式以单倍体形式生长（图 19-6）。复制后的染色体分布于两个非对称的细胞中——祖细胞或母细胞和小一点的芽细胞或子细胞（图 19-6a）。这些细胞可以以两种交配型——a 型和 α 型存在，详见第 10 章和第 17 章。

母细胞和子细胞会表现出不同的交配型，这种差异来自于交配型转换（mating-type switching）过程。母细胞在出芽后能转换交配型，例如，a 细胞产生 a 子细胞，但

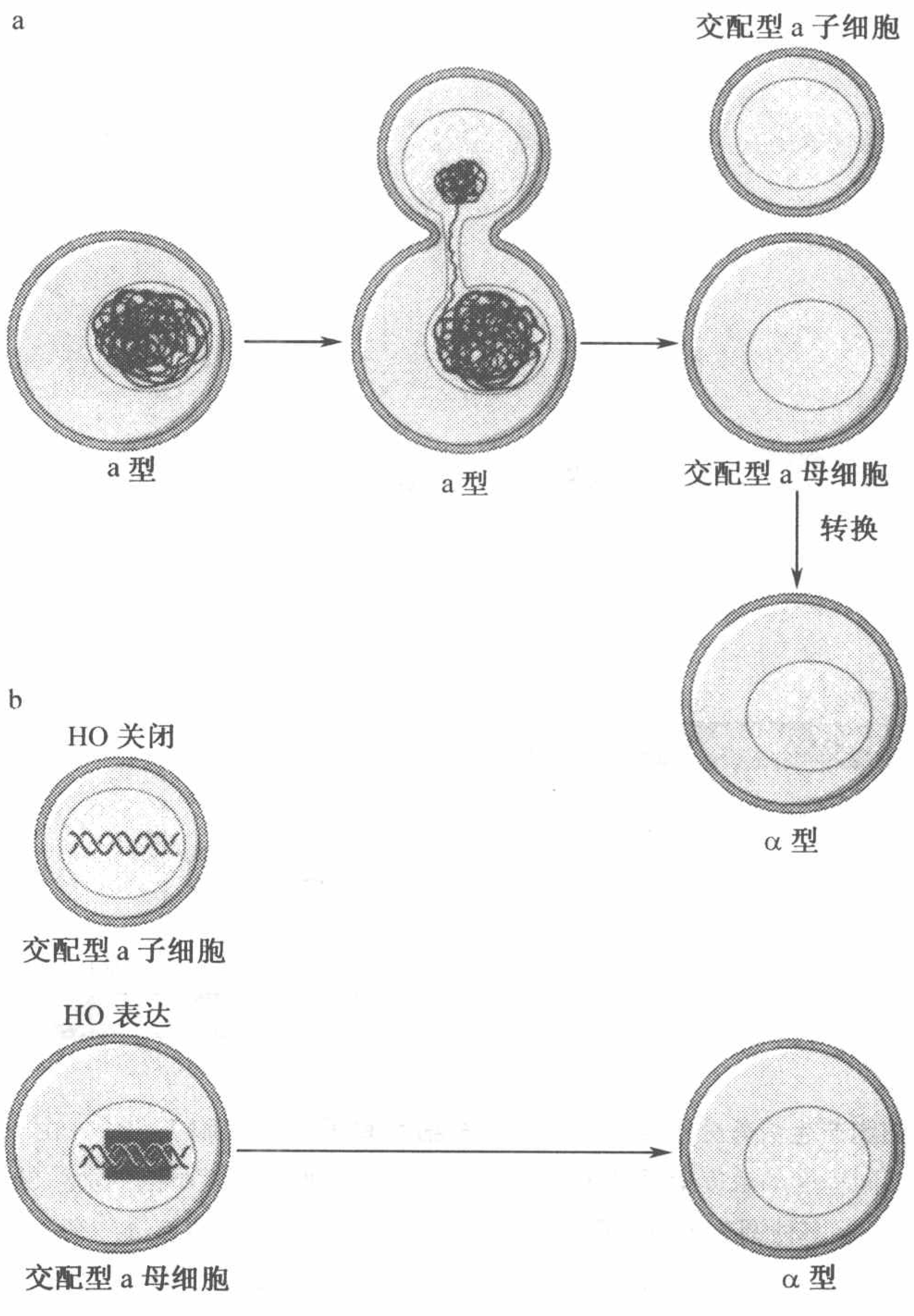


图 19-6　单倍体酵母细胞交配型 a 经过出芽产生母细胞和较小的子细胞。 a. 起始时，两个细胞都是交配型 a，但有时母细胞可以转换为 α 型。b. 子细胞由于具有定位的 Ash1 转录抑制子阻止了 *HO* 基因的表达而不能进行转换。母细胞则不然，由于缺少 *ash1* 基因表达从而可以表达 *HO* 基因。

随之就转换为 α 交配型（图 19-6b）。

转换过程由 *HO* 基因产物控制。从第 10 章我们得知，HO 蛋白是一种序列特异性内切酶。HO 通过对两个沉默的交配型盒之一制造双链断裂来引发交配型位点的基因转换。在第 17 章我们见到了 HO 是如何在母细胞中活化的，在子细胞中由于抑制子 Ash1 的选择性表达导致 *HO* 基因的不表达（图 19-7），这也是子细胞不能进行交配型转换的原因所在。基因 *ash1* 出芽前在母细胞中转录，但编码 RNA 通过以下过程在子细胞中定位。在出芽过程中，*ash1* mRNA 附着在微管的生长端。多种蛋白质起到适配器的作用，其一端结合 *ash1* mRNA 的 3′-UTR，另一端结合微管。微管从母细胞的细胞核延伸到出芽部位，由此方式将 *ash1* mRNA 转运至子细胞。在定位到子细胞后，*ash1* mRNA翻译为抑制子蛋白，该蛋白质与 *HO* 基因结合并抑制其转录。*HO* 基因在子细胞的沉默阻止了子细胞交配型的转换。

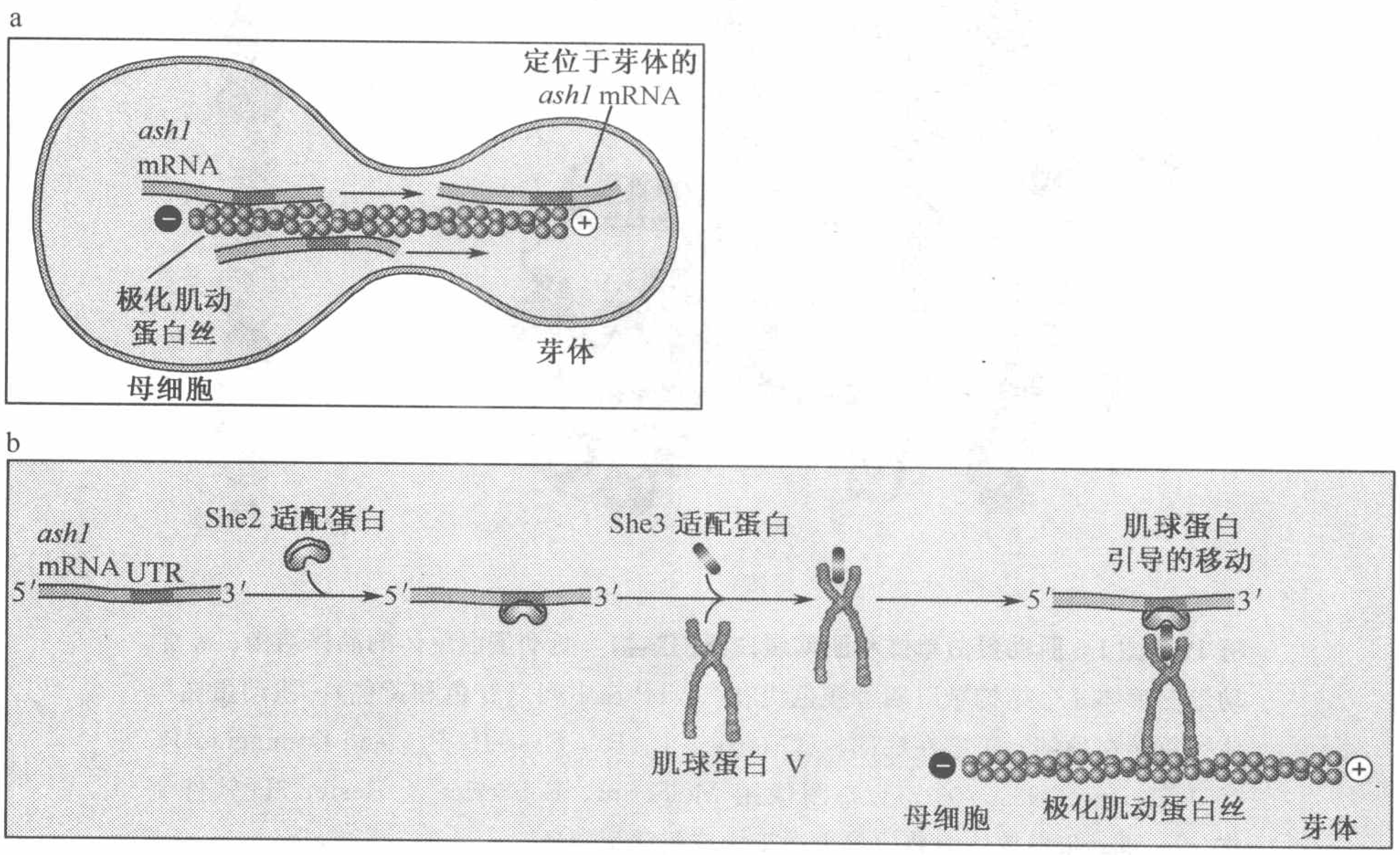


图 19-7　*ash1* mRNA 在出芽过程中的定位。 a. *ash1* 基因在出芽过程中于母细胞中转录。其编码 mRNA 通过带极性的肌动蛋白丝滑入芽体，并从蛋白质丝的负端延伸到生长中的正端。b. *ash1* mRNA 的转运依赖 She2 和 She3 适配蛋白对 3′-UTR 特殊序列的结合。这些适配蛋白结合肌球蛋白，后者携带着 *ash1* mRNA 在肌动蛋白上一起移动。（来源：Adapted from Alberts B. et al. 2002. Molecular biology of the cell，4th edition，p. 971，f. 16-84，part a. Reproduced by permission of Routledge/Taylor & Francis Books，Inc.）

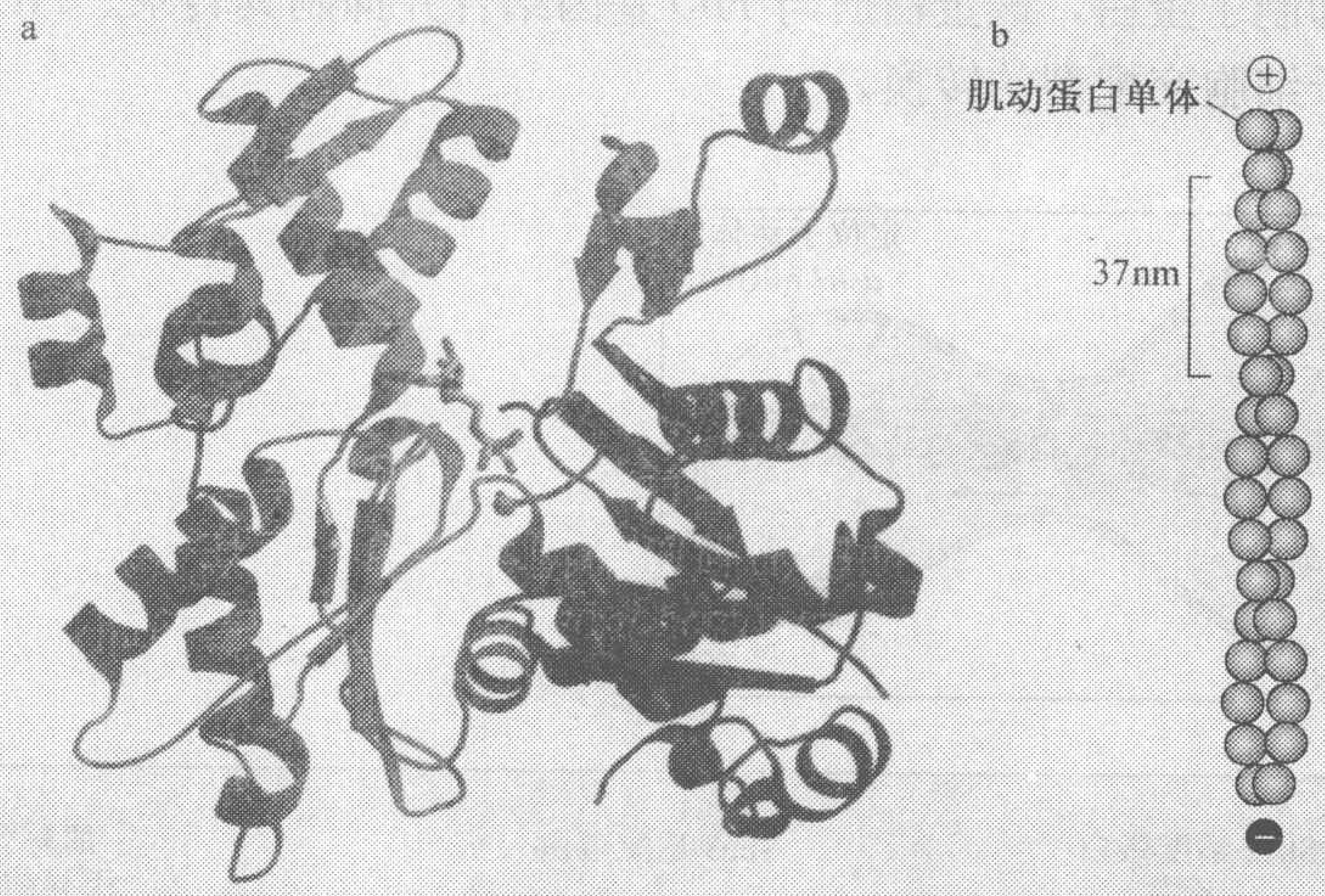
本章后半部分会介绍 mRNA 定位在果蝇发育过程中的作用。同样，这种定位也是通过适配蛋白对 3′-UTR 序列的结合来介导的（框 19-2）。

从酵母的交配型开关研究中发现的第二条规律也同样在果蝇发育中适用：广泛分布的激活子与局部存在的抑制子相互作用，在每个特定细胞中建立起精确的表达模式。在酵母中，SWI5 蛋白质负责 *HO* 基因的激活表达（见第 17 章）。这个激活子在出芽时的母细胞和子细胞中都存在，但其打开 *HO* 的能力仅限于母细胞中，因为在子细胞中有

抑制子 Ash1 的存在。换句话说，尽管有 *SWI5* 基因的存在，但 Ash1 使子细胞的 *HO* 基因处于关闭状态。

框 19-2 细胞骨架：非对称和生长

细胞骨架由 3 种细丝构成：中间丝、肌动蛋白丝和微管。肌动蛋白丝和微管用于各种细胞包括芽孢酵母和果蝇的特异 mRNA 的定位。肌动蛋白丝由肌动蛋白聚合体构成。肌动蛋白聚合体由两条平行螺旋构成，螺旋每周为 37nm。聚合体中的肌动蛋白单体同向排列，因此肌动蛋白丝具有极性。正端较负端生长更快，因而，定位的 mRNA 沿正端移动（框 19-2 图 1）。



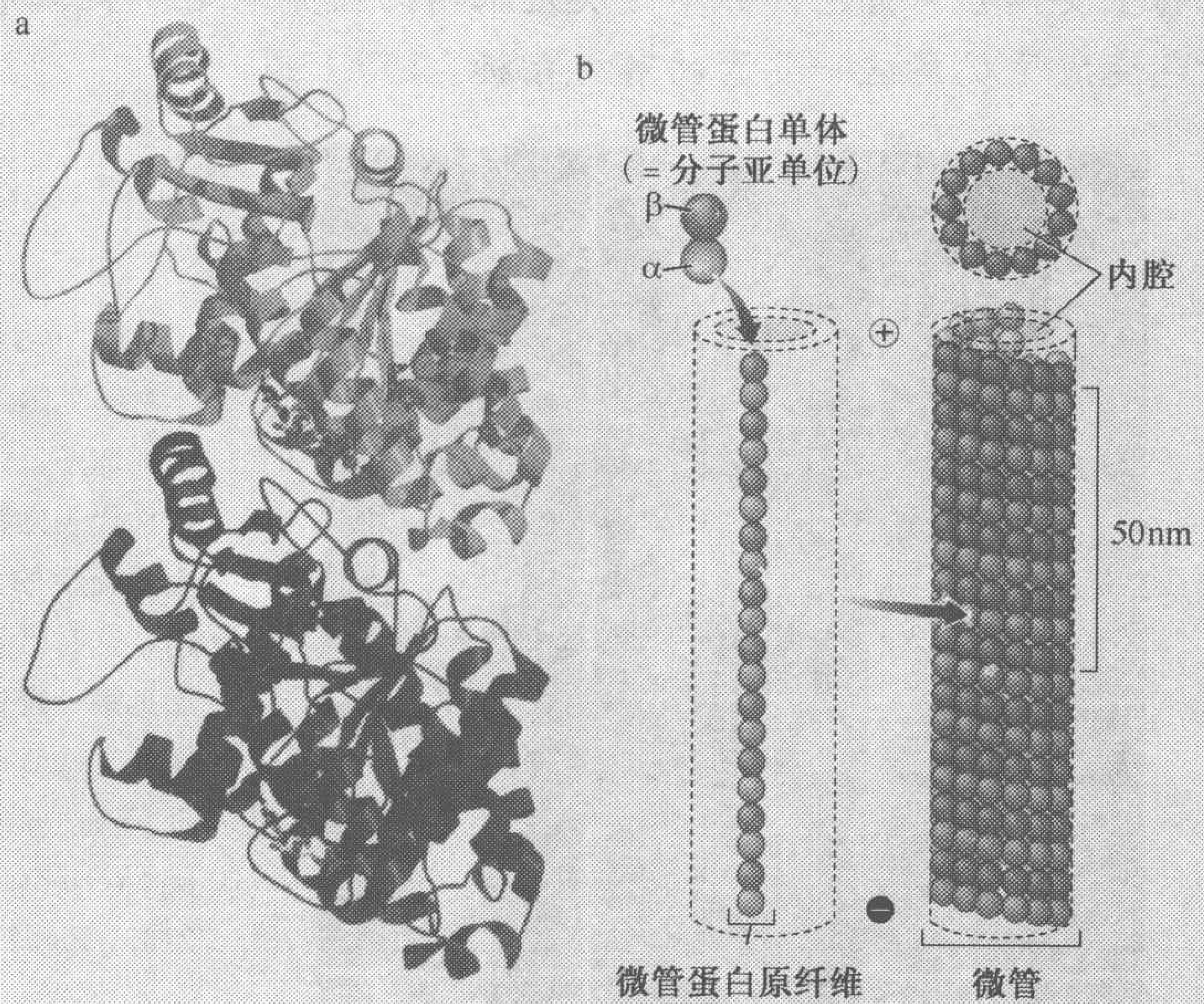
框 19-2 图 1 肌动蛋白单体和肌动蛋白丝的结构。肌动蛋白单体的晶体结构。a. 肌动蛋白单体 4 个结构域以四种颜色代表，ATP 在中间（红色和黄色），肌动蛋白单体的负端在顶部，正端在底部。（Otterbein L. R.，Graceffa P.，and Dominguez R. 2001，Science 293：708-711.）图像用 MolScript、BobScript 及 Raster 3D 软件制作。b. 肌动蛋白单体组装成单个螺旋构成肌动蛋白丝。

微管是由 α 亚单位和 β 亚单位构成的异二聚体微管蛋白的聚合物组成。微管蛋白形成延长的、非对称的原丝（protofilament）。每个微管蛋白异二聚体在原丝中同向排列，13 条不同的原丝在一起平行排列形成柱状的微管。与肌动蛋白丝相似，微管也具有极性，其正端延展迅速而负端相对稳定（框 19-2 图 2）。

肌动蛋白和微管蛋白都具有酶活性，肌动蛋白水解 ATP 为 ADP，而肌动蛋白水解 GTP 为 GDP。这些酶活性启动了肌动蛋白丝和微管的生长或踏车样运动。通常，负端的肌动蛋白或者微管蛋白亚单位介导 ATP 或 GTP 水解并因此变得不稳定而掉下来。相反，正端的亚单位并不水解 ATP 或 GTP，因此相对稳定。

肌动蛋白丝和微管在正端的定向生长依赖于与细胞骨架结合的各种蛋白质。其中一种叫做抑制蛋白（profilin），它与肌动蛋白单体结合促进其在生长正端的整合。其他蛋白质也被证明具有促进微管蛋白原丝从正端生长的作用。

框 19-2 图 2 微管蛋白单体和微管蛋白丝。a. 微管蛋白单体的晶体结构展示了α亚单位（青色）和β亚单位（紫色）。两种亚单位中的 GTP 显现为红色。(Lowe J., Li H., Downing K. H., and Nogales E. 2001, J. Mol. Biol. 313: 1045-1057.) 图像用 MolScript、BobScript 及 Raster 3D 软件制作。b. 微管蛋白的原纤维是由单体按照同一方向排列组成。



定位的 mRNA 启动海鞘胚胎肌肉的分化

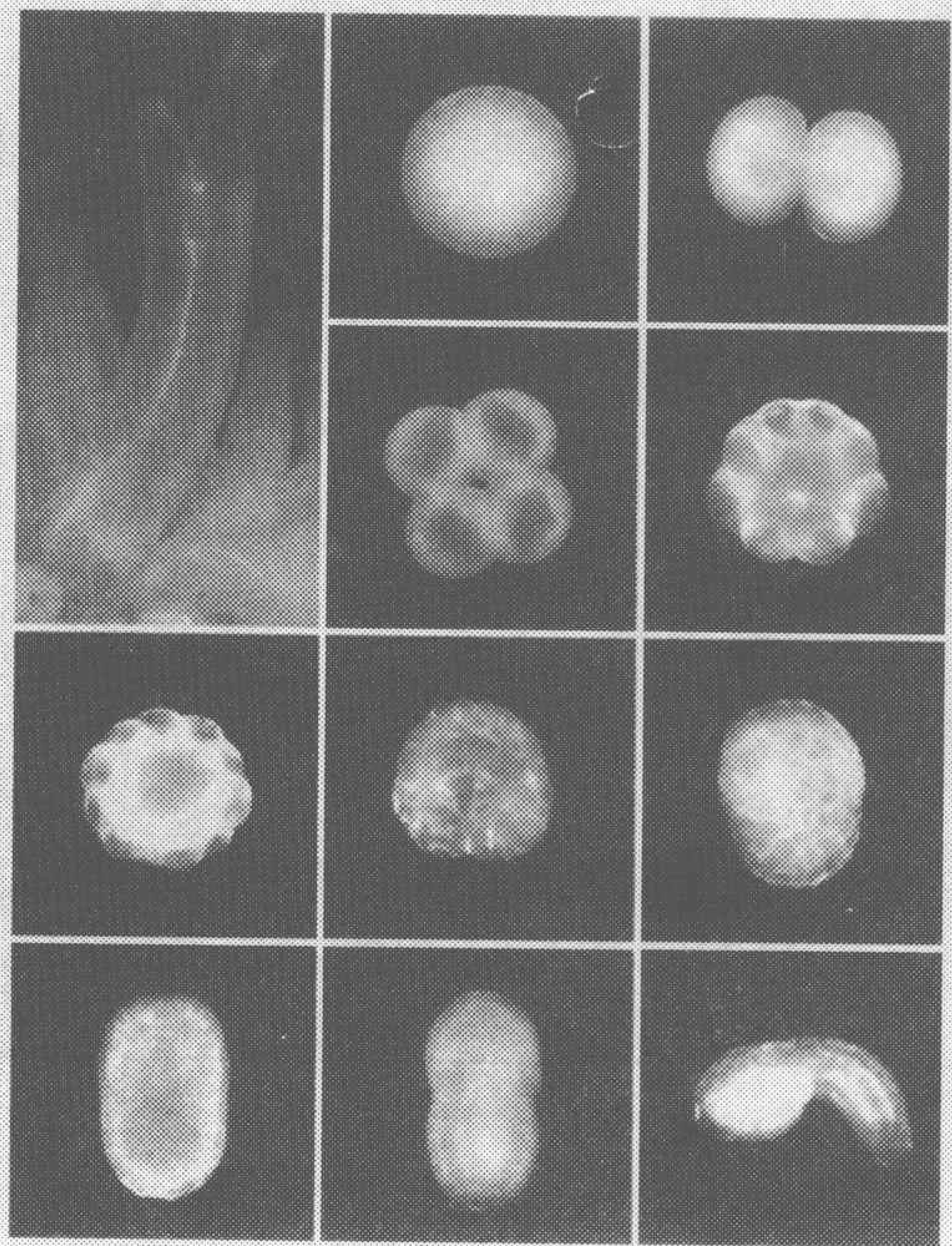
定位后的 mRNA 在相同遗传背景的细胞中能建立差异的基因表达。正如酵母子细胞的命运受制于 *ash1* 基因的遗传一样，发育胚胎细胞可通过遗传定位的 mRNA 而被引导经历特定的发育途径（框 19-3）。

在海鞘肌肉的发育过程中，确定细胞形成肌肉的主要因素为叫做 Macho-1 的调节蛋白。Macho-1 mRNA 在未受精卵的细胞质中均匀分布，但在受精后就局限在卵的植物端（底端）（图 19-8）。最终，在八胚期只有两个细胞遗传了这个 mRNA，这些细胞最终形成尾肌。

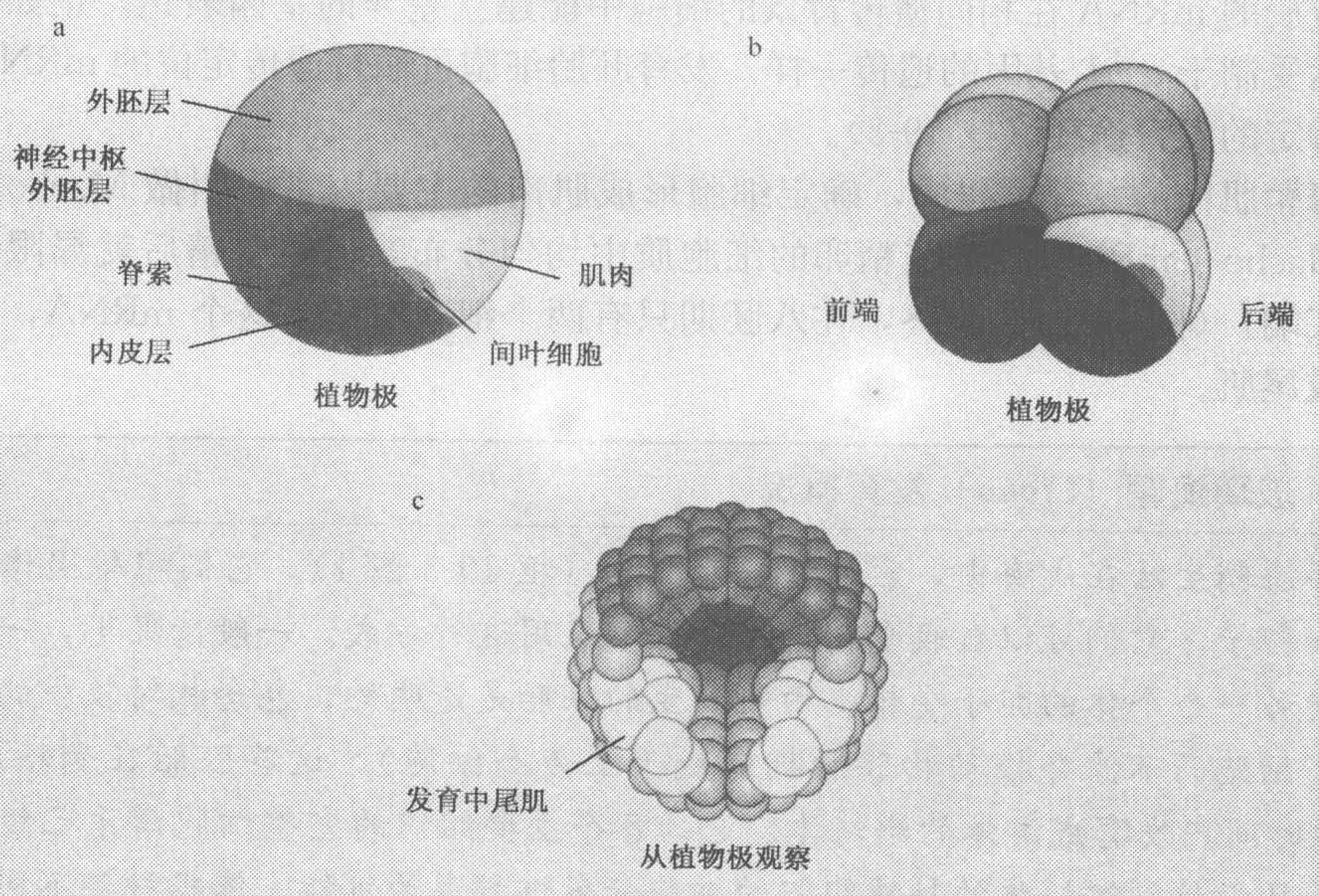
框 19-3 玻璃海鞘（*Ciona*）发育概况

成年海鞘生活在浅海中，靠过滤就地取食（框 19-3 图 1）。它们雌雄同体，同时拥有精子和卵子，它们可以自我受精，但通常不采用这种方式。一般情况下，一个个体的精子可使另一个个体的卵子受精。由此产生的胚胎是透明的，包含相对较少的细胞（只有数百个细胞而不像脊椎动物胚胎拥有成千上万个细胞）。这些胚胎在受精后的 18～24h 的短时间内发育成游泳的蝌蚪体。它的各个主要组织的完整细胞谱系已经清楚，这使得阐明从受精到蝌蚪体的特殊组织的细胞分裂过程成为可能。依据种类不同，蝌蚪体尾部包含 36～40 个肌肉细胞，其谱系可追溯到受精卵。海鞘尾肌代表了一百年前观察到的动物胚胎的第一批谱系。海鞘类非受精卵的黄颜色使这样的观察变得容易。黄色起初分布于整个卵子但在受精后就集中到植物端（底部）（框 19-3 图 2）。定位的黄色素 8 细胞胚胎期或卵母细胞（blastomere）只有其中两个细胞遗传，这两个细胞变成蝌蚪体

尾肌的主要部分。黄色素并不是决定肌肉的要素或者确定细胞是否发育成肌肉的程序，但黄色变成了观察这一要素的可视指标。



框 19-3 图 1 玻璃海鞘生活周期。成年海鞘见图的左上角。橘黄色物质为发育中的卵子，白色物质为精子。以后的照片依次代表逐渐成熟的胚胎。图的第三层代表原肠胚的形成。右下角可看到幼蝌蚪体，这时已是受精后 12～14h（单细胞胚胎见第一排中间图）。（来源：Reproduced from Dehal et al. 2002. The draft genome of *Ciona intestinalis*：Insight into chordate and vertebrate origins. Science 298：2157-2167，Fig. 2，p. 2158.）



框 19-3 图 2 海鞘类的早期胚胎分裂。受精的单细胞胚胎具有多个控制胚胎发育成不同组织的定位要素。例如，黄色代表遗传细胞会形成尾肌，红色代表遗传细胞会形成内胚层或肠。（来源：Redrawn form Gilbert S. E. 1997. Developmental Biology，5th edition，p. 179，Fig5. 17. Copyright © 1997 Sinauer Associates. Used with permission.）

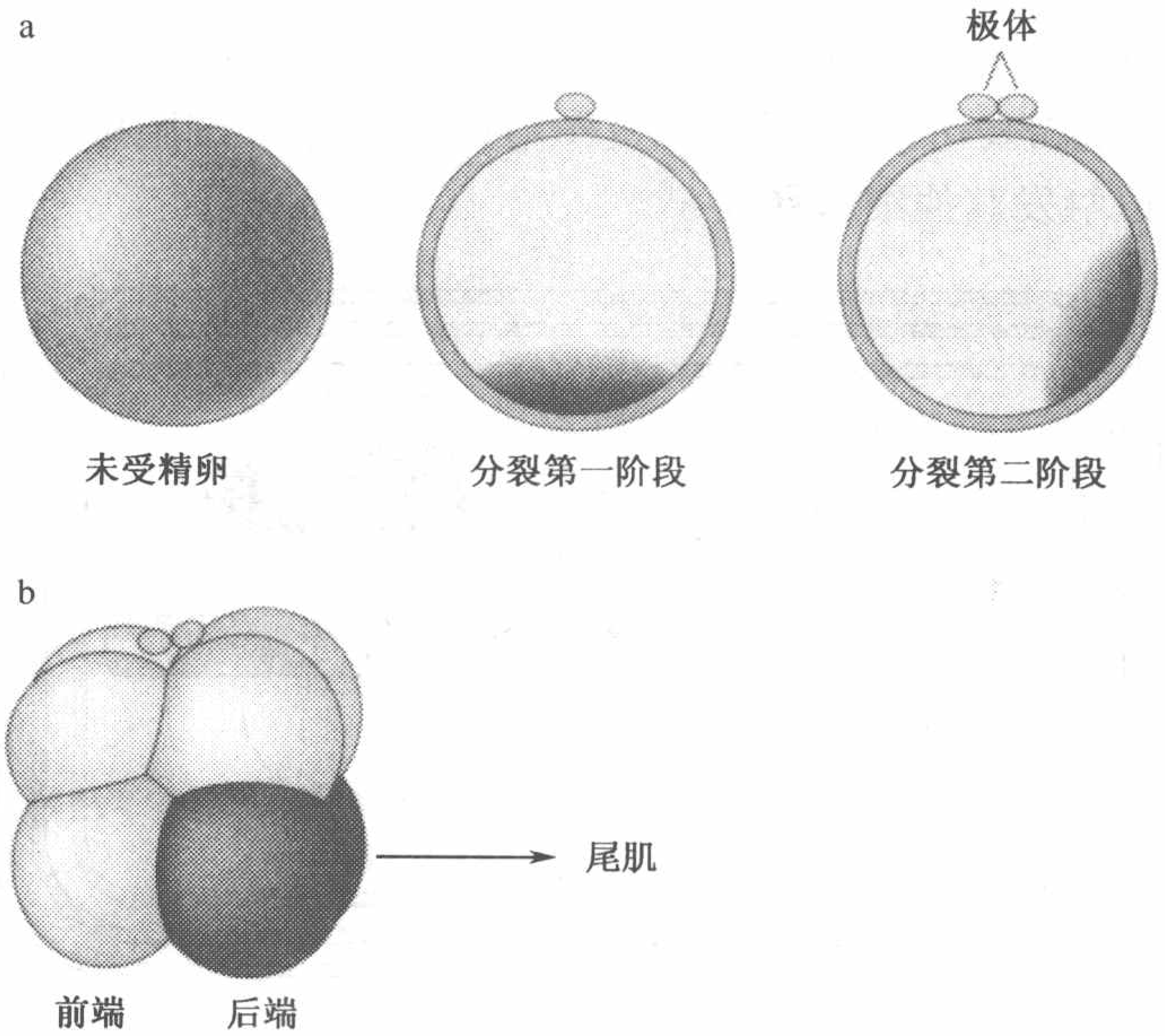


图 19-8 Macho-1 mRNA 在受精卵的定位。 a. Macho-1 在未受精卵的细胞质中均匀分布。受精后卵细胞经过非对称分裂形成极体（顶端）。此时，Macho-1 mRNA 局限于底部。很快，在单细胞胚胎分裂前 Macho-1 mRNA 重新定位。这个过程发生在卵细胞的第二次非对称减数分裂期间。Macho-1 mRNA 分布在单细胞胚胎的一角，此角最终变成 B4.1 卵母细胞（Blastomere）。这些细胞将分化成尾肌。（来源：a. Adapted from Nishida H. and Sawada K. 2001. macho-1 encodes a localized mRNA in ascidian eggs. Nature 409：725，Fig. 1 c，d，e only. Copyright© 2001 Nature Publishing Group. Used with permission.）

Macho-1 mRNA 编码 DNA 结合蛋白——锌指蛋白，在活化肌肉特异基因，如肌动蛋白和肌球蛋白的转录中起作用。因此，这些基因只在肌肉中表达，因为 Macho-1 只在肌肉表达。在本章的第二部分，我们将看到定位 mRNA 合成的调节蛋白如何活化或者抑制基因表达，控制不同类型细胞的形成。

细胞-细胞接触在成孢子细菌激活差异基因表达中的作用

第二种建立差异基因表达的策略是细胞-细胞接触。我们以较为简单的枯草芽孢杆菌为例。在恶劣条件下，枯草芽孢杆菌形成孢子，这个过程的第一步是在孢子前体孢子囊的一端生成隔膜，通过隔膜形成两个不同大小的细胞，小的叫前孢子（forespore），并最终形成孢子；大的叫母细胞，协助孢子的形成（图 19-9）。如下所见，前孢子影响紧邻的母细胞的基因表达。

前孢子拥有活化的特殊的 σ 因子，σ^F，它在母细胞不具有活性。在第 16 章中，我们看到 σ 因子是如何与 RNA 聚合酶结合并选择特定的表达靶启动子的。σ^F 激活 *spo* Ⅱ *R* 基因，后者编码一种分泌编码蛋白 SpoⅡR。SpoⅡR 分泌到母细胞和前孢子的膜间隙，在那里使母细胞的 σ^E 原水解。σ^E 原是 σ^E 的非活性前体，σ^E 原含有 N 端的抑制域，抑制 σ^E 的活性并使其挂在母细胞膜上（图 19-9）。SpoⅡR 引发 N 端多肽的水解，释放出

成熟和活化的 σ^E。σ^E 在母细胞激活一系列与前孢子不同的基因的表达。在这里，SpoⅡR作为信号分子在前孢子和母细胞的界面起作用，通过处理后的活化的 σ^E 引发紧邻的母细胞的差异基因表达。SpoⅡR 诱发作用只能靠细胞-细胞接触，因为所产生的少量的SpoⅡR不足以触发其他细胞释放 σ^E。

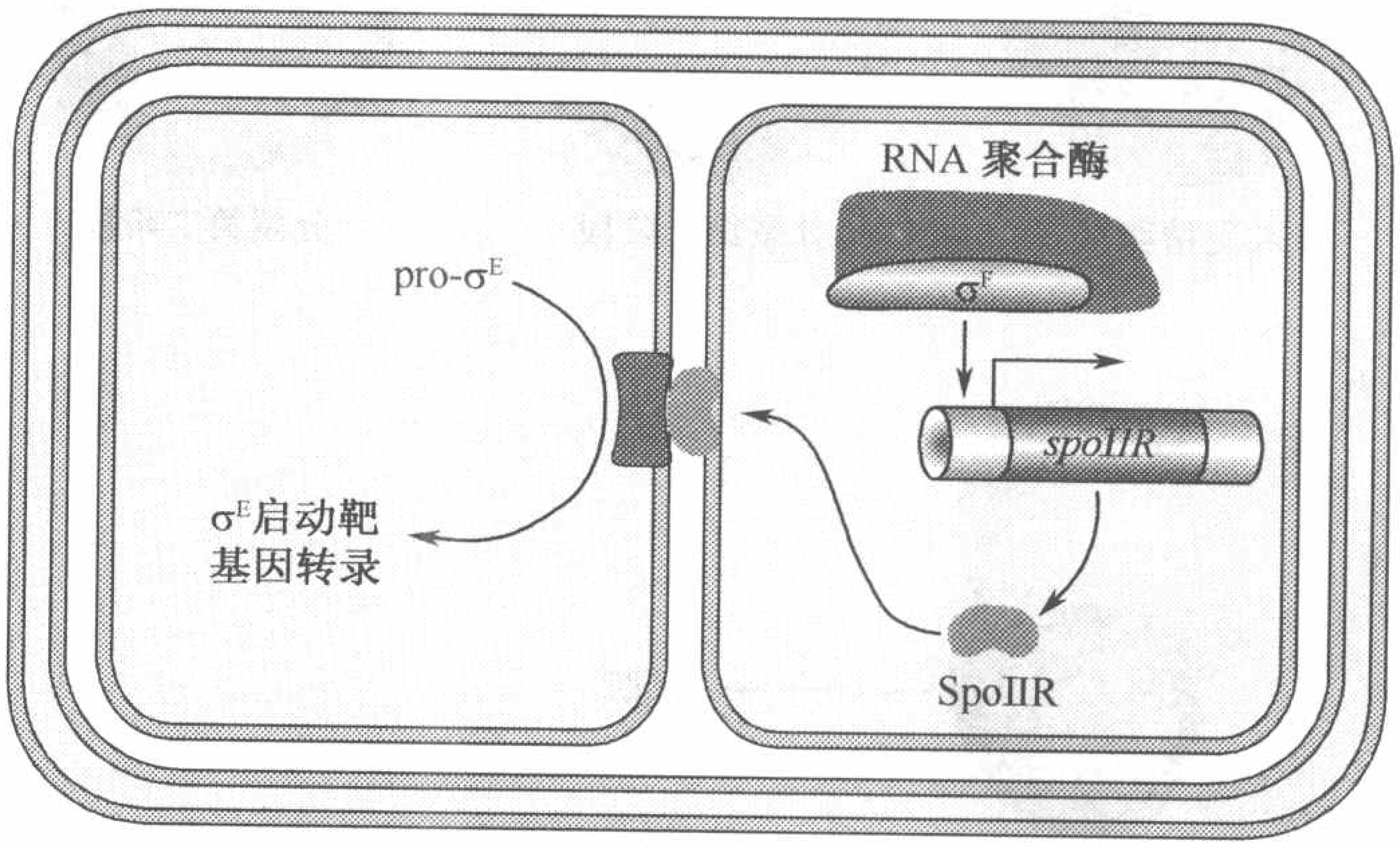


图 19-9 枯草芽孢杆菌母细胞和前孢子的非对称基因活性依赖于不同类型因子的活化。 *spo Ⅱ R* 基因经前孢子的 σ^F 活化，其编码的 SpoⅡR 蛋白与母细胞（左）和前孢子（右）之间的隔膜结合。SpoⅡR 启动母细胞非活性的 σ^E（σ^E 原）的水解使其活化。活化的 σ^E 募集 RNA 聚合酶，激活母细胞特定的基因。（来源：Losick R. and Straiger P. 1996. Sporulation in *Bacillus subtilis*. Ann. Rev. Genet. 30：209，Fig. 3，part a. with permission from the Annual Review of Genetics，Vol. 30. © 1996 by Annual Reviews. www. annualreviews. org.）

昆虫中枢神经的 Notch 信号传导控制皮肤神经调节转换

我们现在转向动物胚胎中的细胞-细胞接触的例子，令人惊异的是它们与枯草芽孢杆菌的过程极为相似。SpoⅡR 引起 σ^E 原水解活化，后者使 RNA 聚合酶与特定基因的启动子序列结合。在下面的例子中，细胞表面的受体受到剪切，其细胞质部分转移到细胞核并在那里结合序列特异的 DNA 结合蛋白进而活化选定基因的转录。

在这个例子里，需要先对昆虫腹神经索的发育做个介绍（图 19-10）。昆虫腹神经索的作用与人类脊椎的作用大致相近。它来自于神经性外胚层（neurogenic ectoderm）的细胞。这个组织可细分为两种细胞：一种留在胚胎表面形成腹部外皮，另一种移入胚胎形成腹神经索的神经元（图 19-10a）。形成皮肤或者神经元的决定来自于这两种细胞的信号传导途径。

发育神经元在其表面拥有叫做 Delta 的信号分子，它与皮肤细胞表面的受体 Notch 结合（图 19-10b）。Delta 对 Notch 的活化使得皮肤细胞不能发育为神经元。这个活化过程使 Notch 的胞内部分 $Notch^{IC}$从细胞膜释放进入细胞核，在那里结合 DNA 结合蛋白Su(H)。产生的 Su(H)-$Notch^{IC}$复合体活化编码转录抑制子的基因进而阻断神经元的发育。

Notch 信号传导并不仅仅只是诱导 Su(H)活性蛋白的表达，而是具有启动调节开关的功能。没有信号传导时，Su(H)可以与多个蛋白质结合，包括 Hairless、CtBP 和

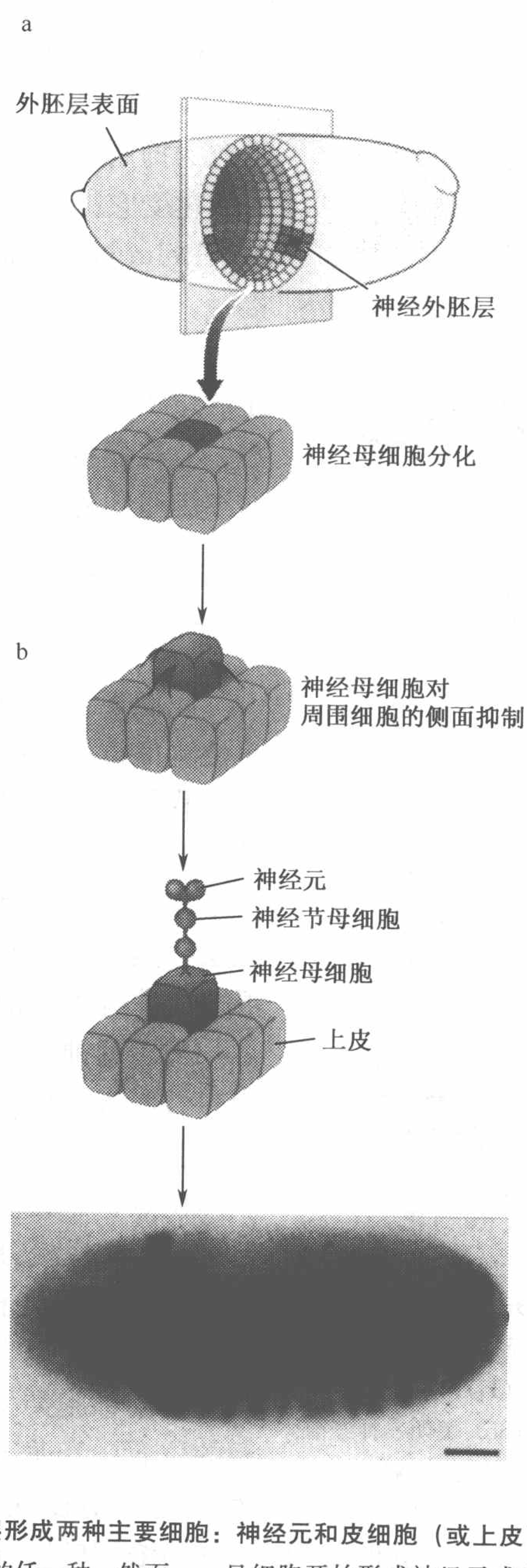


图 19-10　神经外胚层形成两种主要细胞：神经元和皮细胞（或上皮）。 a. 早期神经外胚层能够形成两类细胞的任一种。然而，一旦细胞开始形成神经元或者神经母细胞（neuroblast）（网格中间的黑色细胞），它抑制所有与它直接接触的细胞。b. 这种抑制造成大部分这些细胞保留在胚胎的表面形成皮细胞。相反，发育中的神经元移入胚胎腔形成神经元。（来源：Photo from Skeath J. B. and Carroll S. B. 1992. Regulation of proneural gene expression and cell fate during neuroblast segregation in the *Drosophila* embryo. Development 114：939-946.）

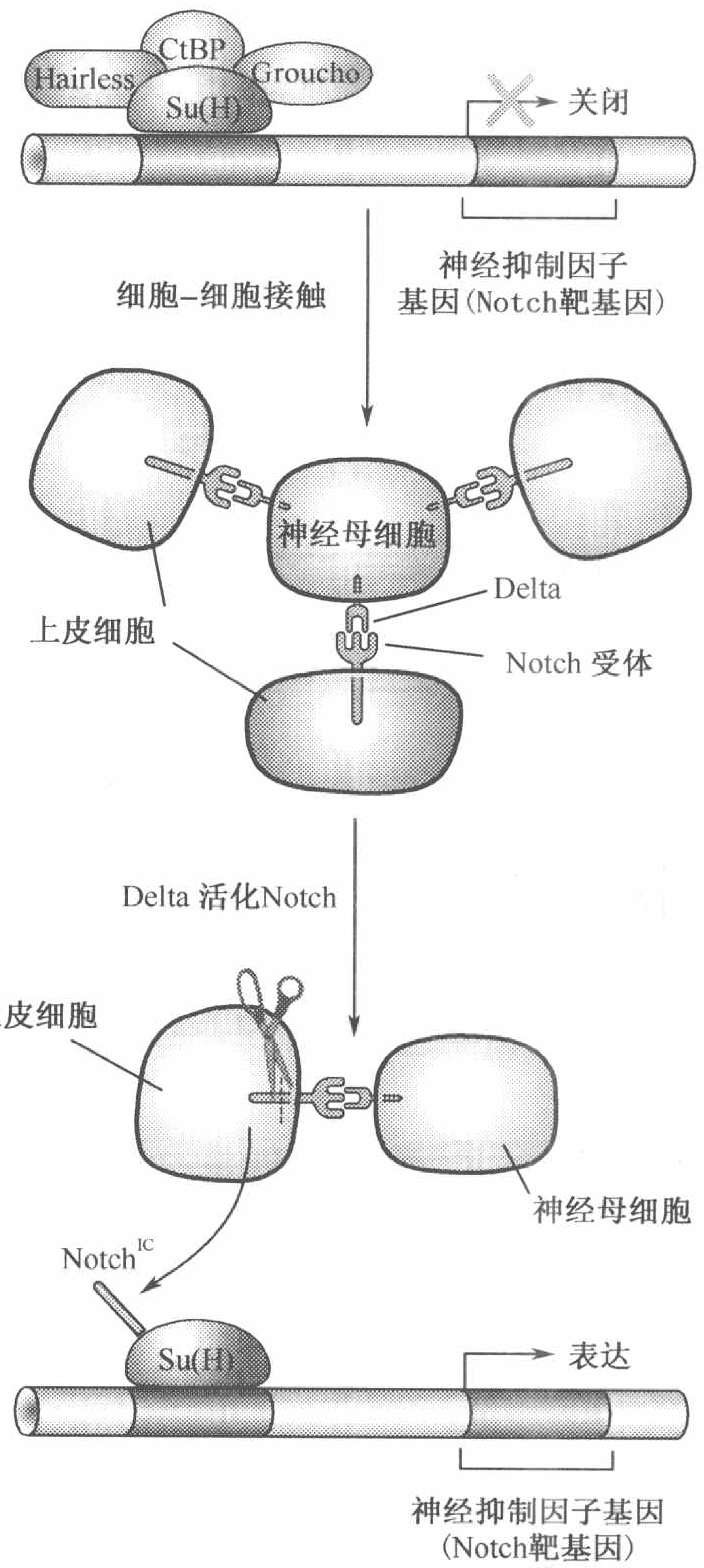


图 19-11 Notch-Su(H)调节开关。发育的神经元（神经元前体细胞）并没有神经抑制子的基因表达。这些基因被 DNA 结合蛋白 Su(H)以及相关抑制子蛋白（Hairless、CtBP、Groucho）关闭。神经前体细胞表达一种膜结合 Delta 信号分子。Delta 与相邻细胞的 Notch 受体结合，Delta-Notch 相互作用造成该细胞的 Notch 受体的活化，进一步分化变成上皮。活化的 Notch 受体经细胞蛋白水解酶剪切，其胞内部分释放进入细胞核。这部分的 Notch 受体使 Su(H)调节蛋白起到活化子而不是抑制子的作用。因此，神经抑制子基因在上皮细胞活化使其不能发育成神经元。

Groucho（图 19-11）。Su(H)与这些蛋白质的任何一个的结合都可以抑制 Notch 的靶基因。$Notch^{IC}$进入细胞核后取代了与 Su(H)结合的抑制因子，使其恢复活性，Su(H)转而活化受它抑制的基因。

Delta-Notch 信号传导依赖于细胞-细胞接触。提供 Delta 的细胞（神经元前体）必须与含有 Notch 的细胞（上皮）直接接触以便活化 Notch 信号传导，抑制神经元分化。下一节我们将给出分泌信号分子如何影响远端细胞基因表达的例子。

Sonic Hedgehog 成形素浓度梯度控制脊椎动物不同神经元的形成

我们现在转向长程信号分子成形素，它将位置信息传递给发育的器官。在这个例子中，我们将继续讨论神经分化，但涉及脊椎动物的神经管。脊椎动物胚胎中，在一定阶段未来背部的所有的细胞都会排列在背部外胚层，然后以协调的方式依次进入胚胎内部形成成体脊椎的前体——神经管。

最靠近神经管腹部的细胞形成特殊的结构叫做底板（floorplate）（图 19-12）。底板是细胞分泌因子 Sonic hedgehog（Shh）的表达位点，它是一种梯度型成形素。

Shh 从底板分泌并在神经管的底部的胞外间质形成浓度梯度（图 19-12a）。神经管内的神经元根据它们接受的 Shh 的数目分化为不同的细胞类型。这取决于它们相对于底板的位置，离底板近的接受高浓度的 Shh；远一些的则接受较低浓度的 Shh。细胞外 Shh 的浓度梯度导致神经管不同细胞的 Shh 受体的不同程度地活化。Shh 浓度梯度至少确立了 4 种细胞的形成（图 19-12b）。

分布于底板附近的细胞接受高浓度的 Shh，其细胞表面活化的 Shh 受体的数目也高。这些细胞接受指令形成叫做 V3 的不同的神经元。远离底板靠近神经管侧面的细胞接受渐次降低的 Shh 蛋白的浓度。因而，

较少数目的活化的 Shh 受体，最终形成运动神经元。较低浓度的 Shh 引导 V2 和 V1 中间神经元的形成（图 19-12b）。

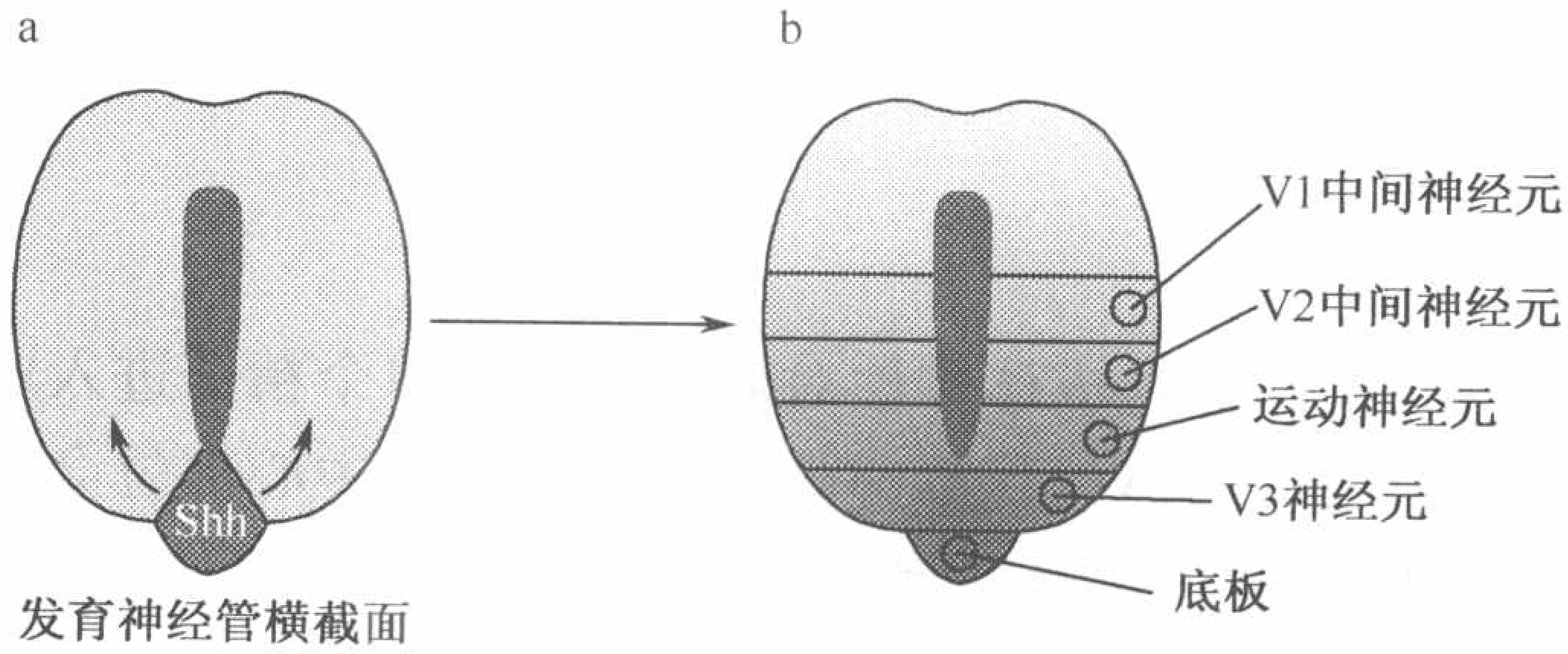


图 19-12　脊椎动物神经管神经元的形成。 a. 分泌信号分子 Sonic hedgehog（Shh）在发育神经管的底板表达（图中底部的棕色圆圈）。Shh 蛋白扩散进入神经管的细胞间质。其在腹部浓度最高，向侧部方向渐次降低。b. Shh 蛋白的阶梯式的分布使得神经管的腹部形成独特的神经细胞类型。它的高和中等浓度的分布导致 V3 和运动神经元的形成。低和更低浓度的分布导致中间神经元的形成。（来源：Adapted from Jessell T. 2000. Neuronal specification in the spinal cord: Inducive signals and transcriptional codes. Nature Rev. Genet. 1: 20-29. Copyright © 2000 Natur Publishing Group Used with permission.）

这些不同程度 Shh 受体的活化又是怎样生成不同的细胞类型的呢？原因是 Shh 受体的活化造成转录活化子 Gli 激活特定的靶基因的表达，Gli 的诱发部分受到它向细胞核调节转运的控制。Shh 结合其细胞表面的受体，使非活化的 Gli 以活化的形式进入细胞核，胞外 Shh 的浓度梯度导致了相应的 Gli 的浓度梯度的形成。也就是说，每一个细胞核内活化的 Gli 的数目与它们离底板的距离有关，越近则浓度越高。

在细胞核内，Gli 以浓度梯度依赖的形式激活基因表达。靠近底板的细胞具有峰值的 Gli 浓度激活促进 V3 神经元分化的基因的表达。稍微低一些的 Gli 浓度则指导运动神经元形成的靶基因的激活。中等和低浓度的 Gli 分别引发 V2 和 V1 中间神经元的形成。下一节，我们会看到 Gli 靶基因调节区 DNA 对 Gli 不同的亲和力在这种 Shh-Gli 靶基因的分化调节中发挥了重要作用。V1 基因受低浓度 Gli 活化，因为它们调节区的序列对 Gli 具有高亲和力。相反，V3 靶基因具有低亲和力的 Gli 识别序列，只受峰值 Shh 信号和 Gli 活化子浓度的活化。这种产生基因表达和细胞分化的多种阈值的调节原则在早期果蝇胚胎中表现得尤为突出。

果蝇胚胎发育的分子生物学

在本章的其余部分，我们集中描述果蝇的胚胎发育。果蝇在这个方面的分子细节较之于任何其他动物胚胎来讲最为完备。本章前面讨论的细胞通信的各种机制以及前面章节讲到的基因调节在这个例子里都会涉及。

定位要素和细胞信号途径都用来建立产生位置信息的调节蛋白的梯度浓度，这个浓

度造成前-后（头-尾）和脊-体（背-腹）体轴的形成。这些调节蛋白活化子或者抑制子控制那些产生胚胎不同的确定区域的基因表达。常见的现象是对复杂调节 DNA（尤其是复杂的增强子）的利用，经过转录活化子和抑制子的交替作用产生基因表达的明显开关效应。

果蝇胚胎发育概述

果蝇生命周期与人类相似：成体雄性对雌性受精。单个精子进入成熟卵子，单倍体精子和卵子的核融合形成二倍体的受精卵核。这个核卵的中心部位会经历一系列几乎同步的分裂。由于没有细胞膜来分隔细胞核，胚胎就变成了所谓的合胞体（syncitium），即单个细胞多个核。下一个系列的分裂中，核开始向卵的周边移动。到达外周后，核完成 3 次分裂，约 6000 个核最终沿着中心的卵黄围成一圈。在受精 2～3h 后，经过 1h 的时间在核之间形成细胞膜。

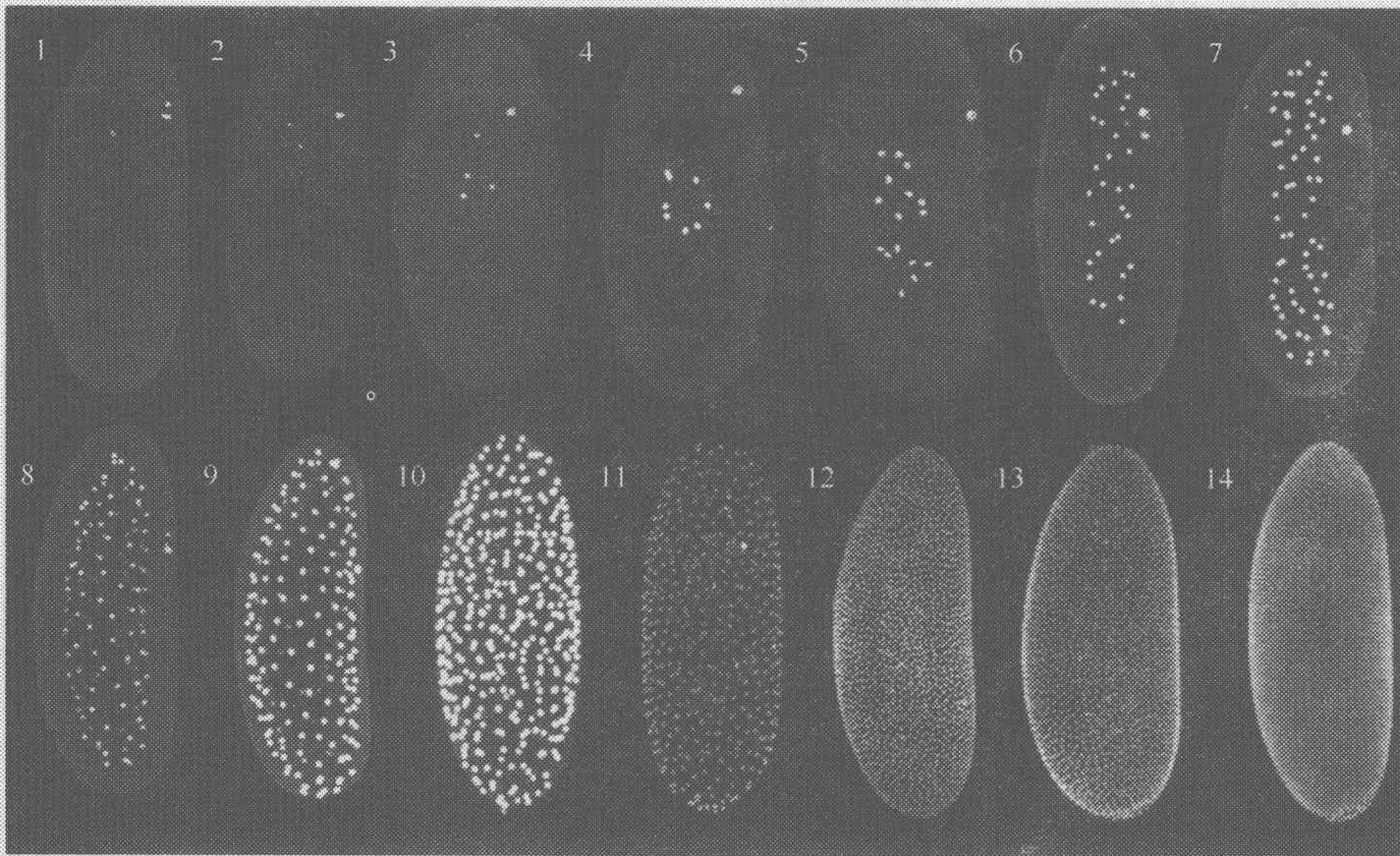
在形成细胞膜之前，核是全能的（totipotent），或者是未予授权的：它们并没有什么特征，仍然能够产生任何细胞类型。一旦细胞化，核只有不可逆地分化进入果蝇的特定组织。框 19-4 叙述了这个过程，以下部分将叙述确立这种剧烈变化过程的分子机制。

框 19-4 果蝇胚胎发育概况

精卵单倍体核融合后，受精卵在中心卵黄区经历一系列 10 轮近乎同步的快速分裂。从核中心粒发散的大的微管阵列将核从中心拉向卵的外周（框 19-4 图 1）。在 8 次分裂后，有 256 个受精卵核开始向外周移动，移动过程中又有两次分裂（框 19-4 图 1，核分裂第 9 轮）。几乎所有这些约 1000 个核进入卵的皮层区（框 19-4 图 1，核分裂第 10 轮）。其余细胞（vitellophage）留在中央区域，在发育中作用不明。

在受精后 90min 时，大部分核到达皮层，获得转录 Pol Ⅱ基因的能力。像许多其他生物如蛙类，似乎有一个“中囊胚转变”阶段，早期的卵母细胞（blastomere）（或者核）在快速有丝分裂时表现出转录沉默。相关机制并不明确，似乎 DNA 在进行高速复制时很难同时维持转录。这些以及其他观察结果提示复制和转录在竞争同样的大分子复合物。由于转录只有在核到达皮层后才能进行，所以猜测外周区域可能含有相关的定位要素。最近的基因表达研究已经揭开了许多皮层的奥秘。例如，分节基因（segmentation gene）*hunchback* 在早期胚胎的前半部分的核中均匀表达，包括进入皮层的和留在卵黄的消黄细胞（vitellophage）的表达。

核到达皮层后又进行了 3 轮分裂（受精后共有 13 轮分裂），产生了 6000 个排列紧密的柱形的核包围着卵黄的包装（框 19-4 图 1，核分裂第 14 轮）。从细节上讲，此时胚胎仍是合胞体，但抗细胞骨架蛋白抗体对早期胚胎组织的染色表明每个核周围呈现密实的网状结构。受精 2～3h 后的 1h 内，胚胎经历了剧烈的细胞化过程，相邻的核之间形成了细胞膜（框 19-4 图 1，核分裂第 14 轮）。受精后 3h，胚胎转化为细胞母胚层（blastoderm），与大部分其他胚胎的“细胞孔壁”一致。

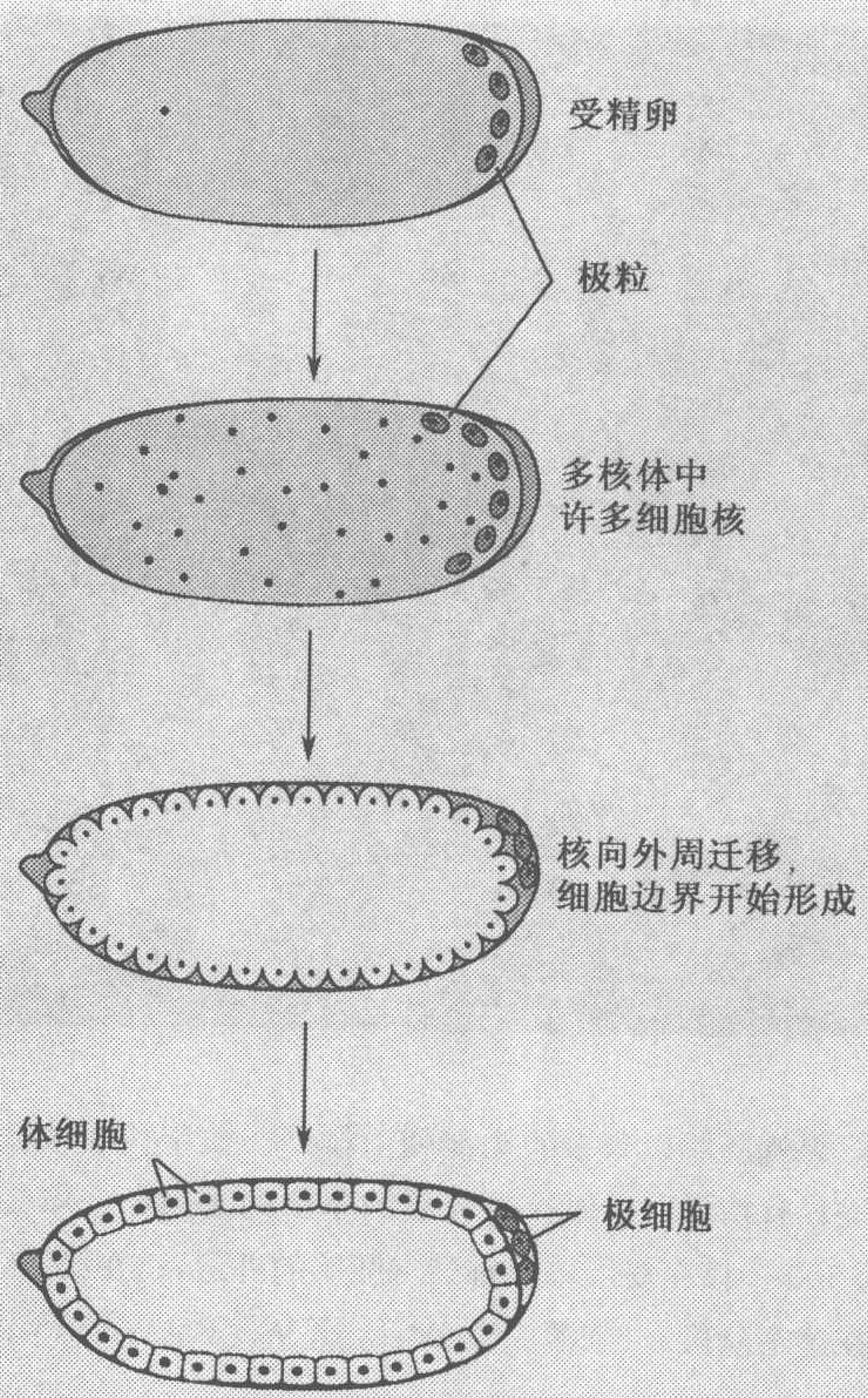


框 19-4 图 1 果蝇胚胎发生。果蝇胚胎以未来的头部朝上排列。数字代表核分裂的数字。胚胎内的细胞核染成白色。例如，在第一阶段只有精子卵子原核融合后的单个的受精卵；第二阶段包含受精卵核的第一次分裂；在第 10 阶段时已有约 500 个核在皮层单层排列（胚胎的外层）；在第 14 轮时已有 6000 多个核单层紧密地列在皮层，细胞化发生在这个阶段。（来源：Courtesy of W. Baker and G. Shubiger.）

传统胚胎学引人注目的精髓之一就是胚胎本身。大部分海生生物的胚胎如海鞘类看起来赏心悦目。而果蝇的胚胎丑陋不堪，聊以慰藉的是它含有的大量的基因表达模式。早期胚胎以各种分子的、组织学的方式为我们详尽地展示了差异基因活性所彰示的不仅仅是细胞命运的特化。进一步来说，第一批展示的基因就有编码控制细胞命运的调节蛋白。因而，分子研究一步到位地揭示了细胞命运特化和确定的神秘过程。

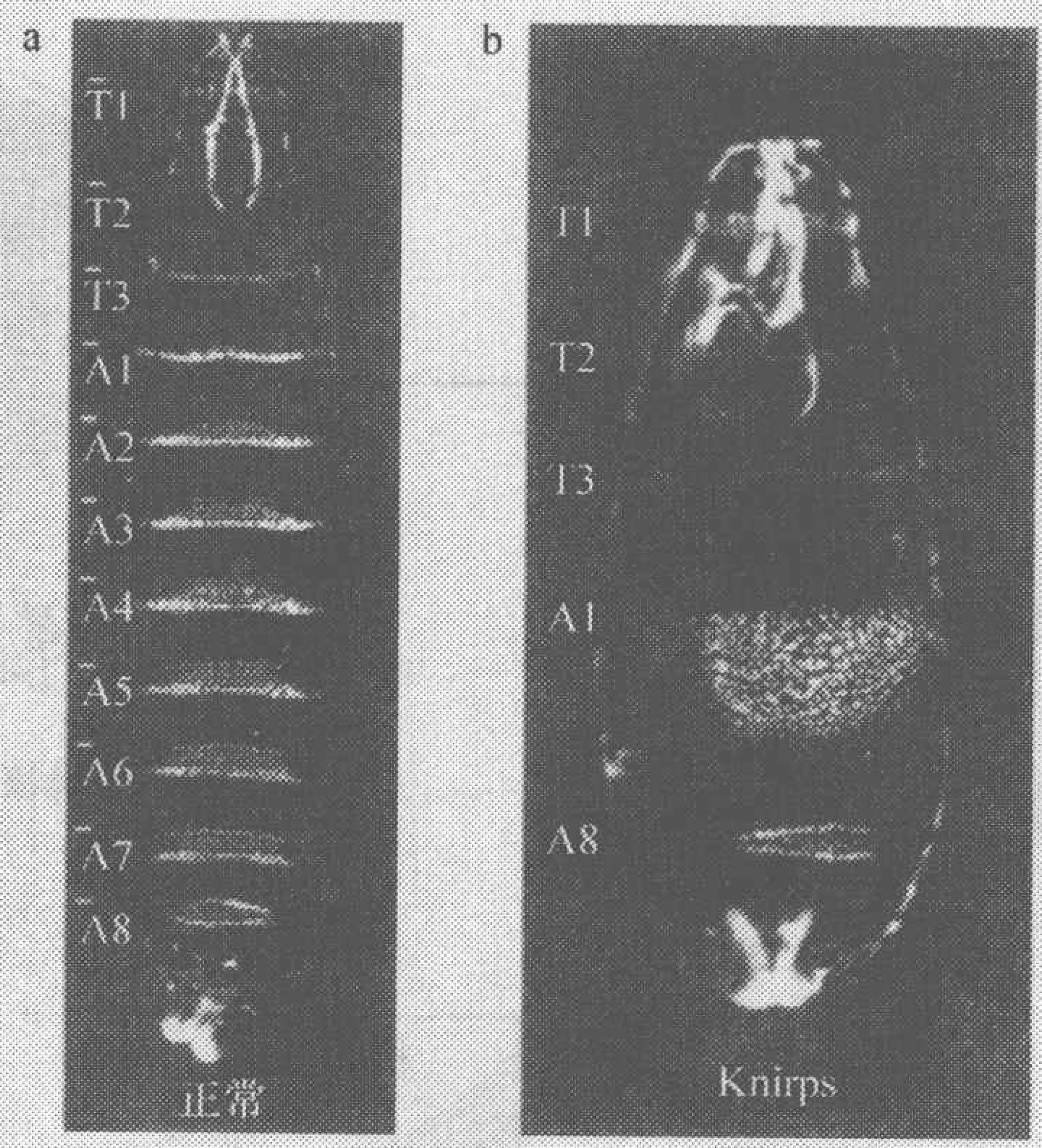
在核进入卵的皮层时，它们是全能的，可以形成成体的任何细胞，但它们的位置这时决定了它们的命运。移到皮层后部的 30 多个核遇到那里的蛋白质组分，如 Oskar，它设定将原细胞变成生殖细胞的程序（框 19-4 图 2）。在一端的细胞质中可能的要素有叫做极粒（polar granules）的核酸蛋白复合体。后部的核在此部位与胚胎的主体断开，产生的极细胞依据卵子的性别分化为精子或卵子。在中部或前部显微注射后端胞质引起那里分化形成大量的极细胞。

没有进入极性胞质的皮层细胞核将形成体细胞组织。这些细胞是全能的，可以形成任何细胞类型。在一个极短的时间内，也许短至 30min，每个核就确立了分化的特殊途径。这个过程发生在细胞化阶段，没有理由证明核之间细胞膜的形成与决定细胞命运有直接关系。在完成细胞形成时不同的核表现出不同的基因转录类型。受精后 3h，细胞已经得到了固定的位置标志，那些胚胎前部的细胞形成成体果蝇的头部结构，而后部的则形成腹部结构。



框 19-4 图 2 生殖细胞的发育。极粒分布在未受精卵的后部，这里包含生殖细胞决定因素和 Nanos mRNA，后者对腹节的发育有重要作用。核（中间的点）向外周移动，进入后部的核分隔为极粒形成极细胞进而形成生殖细胞，其余细胞（体细胞）形成成体果蝇的所有其他组织。[来源：Adapted from Schneiderman H. A. 1976. Insect development. In Symposia of the Royal Entomological Society of London 8：3-34.（ed. P. A. Lawrence）. Copyright © 1976. Reprinted by permission of Blackwell Science.]

由 Eric Wieschaus 和 Christiane Nüsslein-Volhard 进行的系统遗传实验鉴定了 30 多个控制果蝇胚胎类型的分节基因（segmentation gene），涉及对上千个死胚胎的检验。在胚胎发生的中间点，腹皮（上皮）分泌含有细毛或小齿的表皮（cuticle），以死胚小齿类型的破坏为标准可将这些分节基因分为 3 类。缺口基因的突变造成数个紧邻的分节的缺失（框 19-4 图 3），例如，*knirps* 的突变造成腹节 2～7 的缺失（正常有 8 节）；结对基因（pair-rule）突变造成腹节每隔一节的缺失，例如，*even-skipped*（*eve*）基因的突变造成偶数腹节的缺失；最后，腹节极性基因的突变并不改变腹节的数目但会使每一个腹节产生缺陷。正常腹节只在一定区域有小齿，其余部分裸露，腹节极性基因的突变会使这两部分都长小齿。



框 19-4 图 3　正常和突变腹节的暗视野图像。 a. 正常胚胎体节小齿的类型在不同的体节有些许不同（标为 T1～A8 图像）；b. Knirps 突变体（缺口基因 *knirps* 突变）缺失第 2～7 腹节。（来源：Nusslein-Volhard C. and Wieschaus E. 1980. Mutations affecting segment number and polarity in *Drosophila*. Nature 287：795-801. Images courtesy of Eric Wieschaus，Princeton University.）

成形素浓度梯度控制果蝇脊-腹成形

早期果蝇胚胎脊-腹成形受到调节蛋白 Dorsal 的控制，此蛋白质均匀分布于未受精卵的细胞质。受精核转移到卵外周后，Dorsal 蛋白从底部和侧部进入细胞核但仍然保留在细胞质的顶部（图 19-13）。卵细胞核里的 Dorsal 会形成与神经管底部细胞 Gli 相似的浓度梯度。

Dorsal 蛋白对核转运的调节由信号分子 Spätzle 控制。这个信号在未受精卵和外卵壳（outer egg shell）细胞膜之间的胞外间质存在一个浓度梯度。受精后，Spätzle 与细胞膜表面 Toll 受体结合。根据 Spätzle 浓度的不同和合胞体胚胎受体占位程度，Toll 受体的活化程度也不同。Toll 受体在腹区由于 Spätzle 浓度最高因而活化程度也最高，侧面区域的活化程度依次减低。Toll 信号传导造成细胞质 Cactus 抑制因子的降解进而 Dorsal 释放并从细胞质进入细胞核。

某些 Dorsal 靶基因的活化需要 Dorsal 蛋白的峰值浓度，另一些只需要中等或低浓度。这样受精后 2h 内，在经历着细胞化的胚胎中，Dorsal 的浓度梯度就设定了横跨脊-腹轴 3 个基因表达的阈值。这个阈值启动了 3 种各具特点的组织的分化：中胚层、腹神经外胚层和脊神经外胚层（图 19-14）。每个组织继续在成体果蝇形成特定的细胞，中胚层形成飞行肌和内脏，如类似于人类肝脏的脂肪体；腹神经外胚层和脊神经外胚层形成腹神经索的各种细胞。

我们现在考虑 3 个不同的受到高、中、低浓度 *Dorsal* 基因活化的基因：*twist*、*rhomboid*、*sog*。Dorsal 梯度浓度的最高处即细胞核活化最靠近腹部的 18 个细胞的

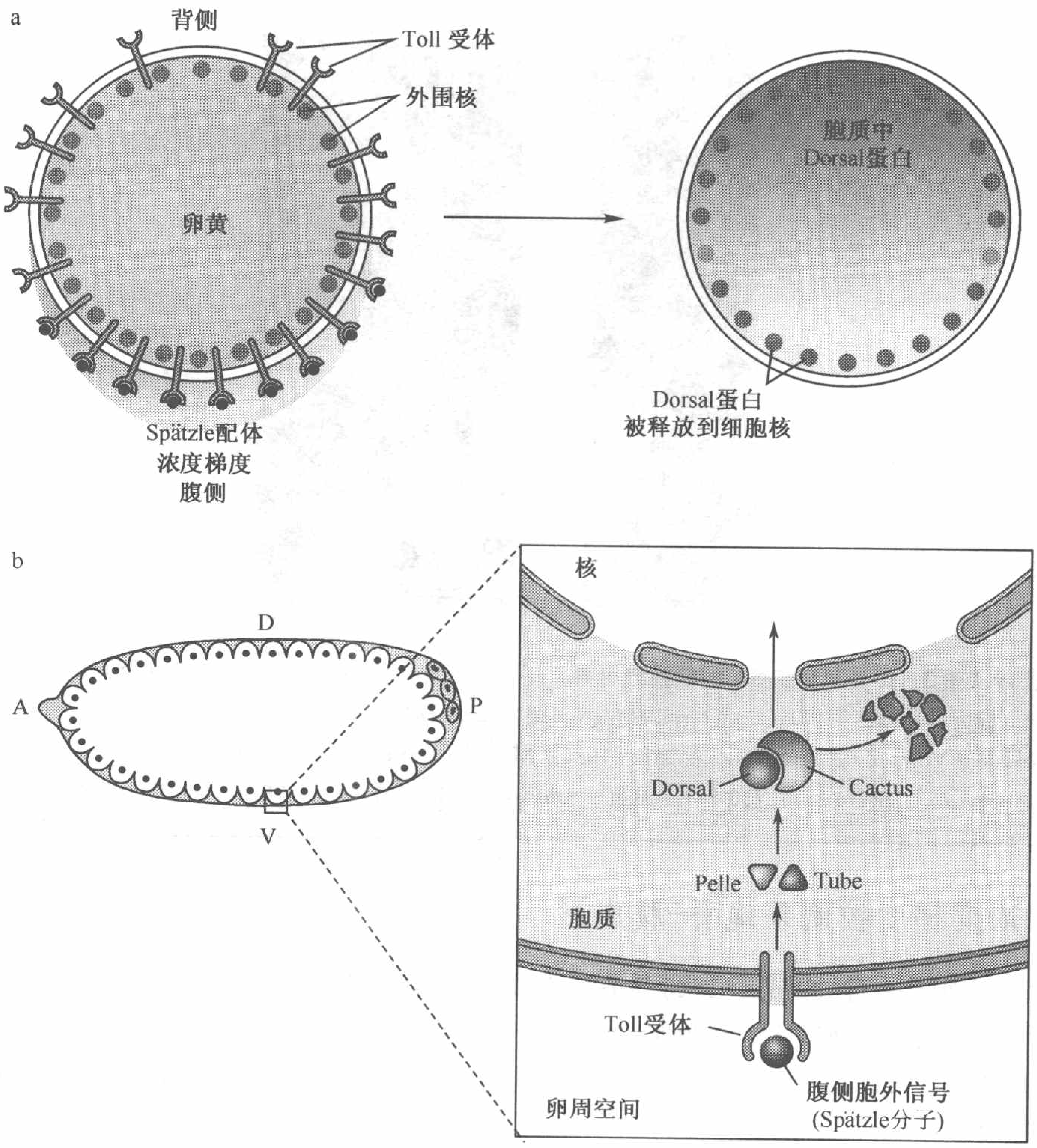


图 19-13 Spätzle-Toll 和 Dorsal 浓度梯度。 a. 圆圈代表早期果蝇胚胎的剖面。Toll 受体均匀分布在前细胞胚胎的细胞膜上。Spätzle 信号分子存在浓度梯度，最靠近腹部处最高，其结果是腹部的活化受体要多于侧部和脊部。Toll 信号传导的浓度梯度形成了 Dorsal 在核内的浓度梯度。b. Toll 信号传导的级联反应。Toll 受体的活化导致细胞质中 Pelle 激酶的活化。Pelle 直接或者间接使 Cactus 磷酸化，后者结合和抑制 Dorsal 蛋白。Cactus 的磷酸化导致它的降解，Dorsal 就被释放并转移进入细胞核。

twist 基因（图 19-14）。胚胎侧部没有 *twist* 基因的活化，那里具有中和低含量的 Dorsal 蛋白。因为 *twist* 基因的 5′端调节 DNA 包含两个低亲和力的 Dorsal 结合位点（图 19-14）。因此，需要高浓度的 Dorsal 蛋白来有效占据这些位点，侧部的 Dorsal 蛋白的浓度不足以进行有效的结合和激活 *twist* 基因的转录。

rhomboid 基因由腹神经外胚层的中等浓度的 Dorsal 蛋白活化。在 *rhomboid* 基因转录起始点上游 1.5kb 处有一段 300bp 的增强子（图 19-15a）。此增强子包含多个 Dorsal 结合位点，多为 *twist* 基因 5′端的低亲和力的调节区，其中至少有一个位点有适当的亲

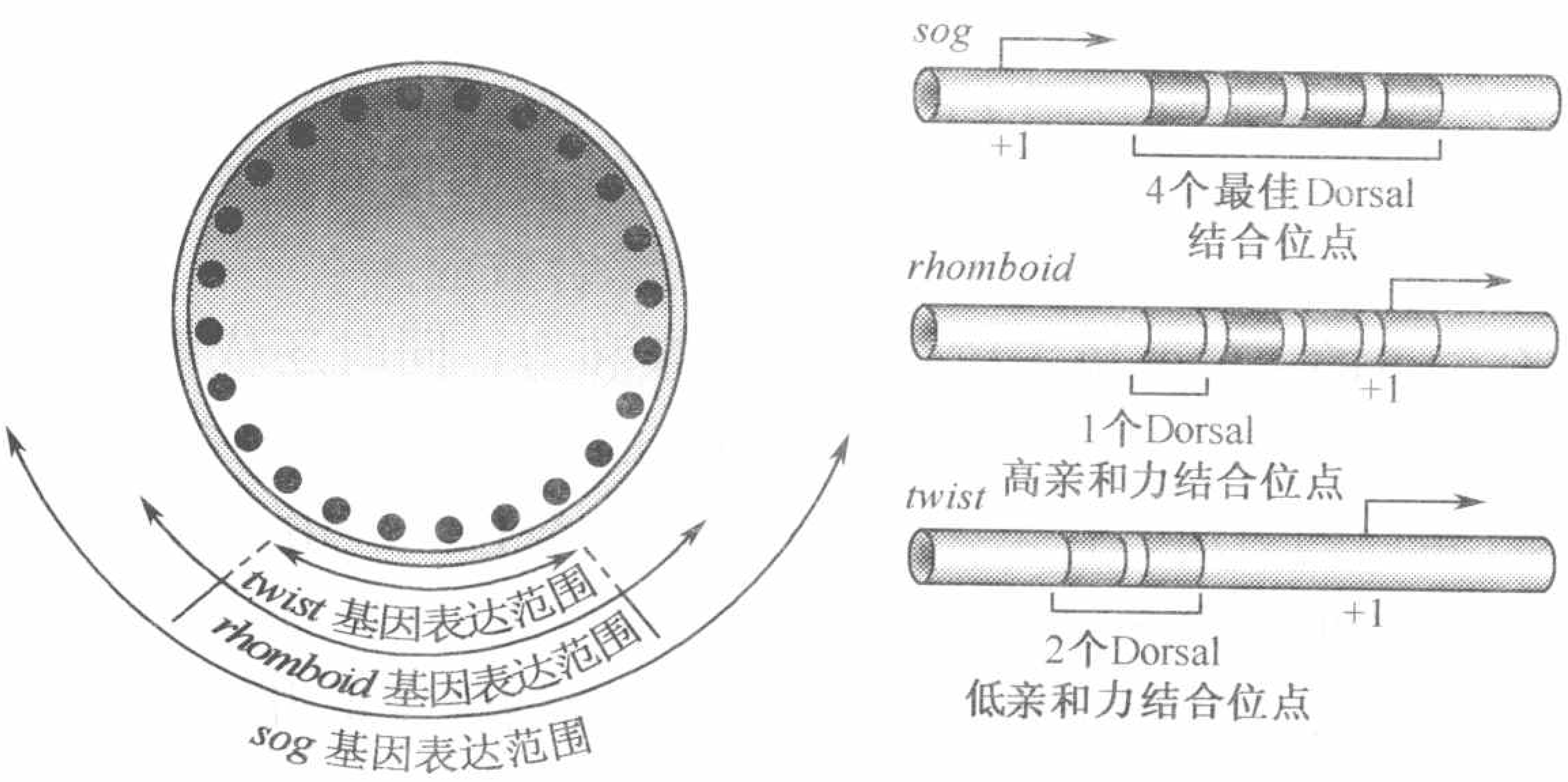


图 19-14　3 种阈值和 3 种调节 DNA。 *twist* 基因 5′端 DNA 含有两个低亲和力的 Dorsal 结合位点，二者的占位只出现在 Dorsal 低度浓度条件下。其结果，*twist* 的表达局限在腹部的核内。*rhomboid* 5′端的增强子包含多个 Dorsal 结合位点，其中只有一个代表了最佳高亲和力 Dorsal 识别序列。低和高亲和力位点的混合存在允许了高和中等浓度的 Dorsal 蛋白激活 *rhomboid* 基因在腹侧部位的表达。最后，*sog* 基因内含子增强子包含 4 个等距离的最佳 Dorsal 结合位点。这样，高、中、低浓度的 Dorsal 蛋白都可以活化 *sog* 基因在果蝇侧面部位的表达。

和力与侧部的中等浓度的 Dorsal 蛋白结合。原则上，*rhomboid* 增强子可以被中胚层的高浓度的和腹神经外胚层的中等浓度的 Dorsal 蛋白激活，但是在中胚层它被转录抑制子 Snail 关闭。Snail 抑制子只在中胚层表达而不在神经外胚层表达。300bp 的 *rhomboid* 增强子同时包含 Snail 抑制子和 Dorsal 活化子的结合位点，这种广泛分布的 Dorsal 活化子和局部分布的 Snail 抑制子的相互作用造成了 *rhomboid* 基因在腹神经外胚层的有限表达。在酵母，我们已经看到局部的 Ash1 抑制子如何阻断子细胞 SWI5 活化子的作用，后面我们还会看到这种方式在果蝇发育中的大量出现。

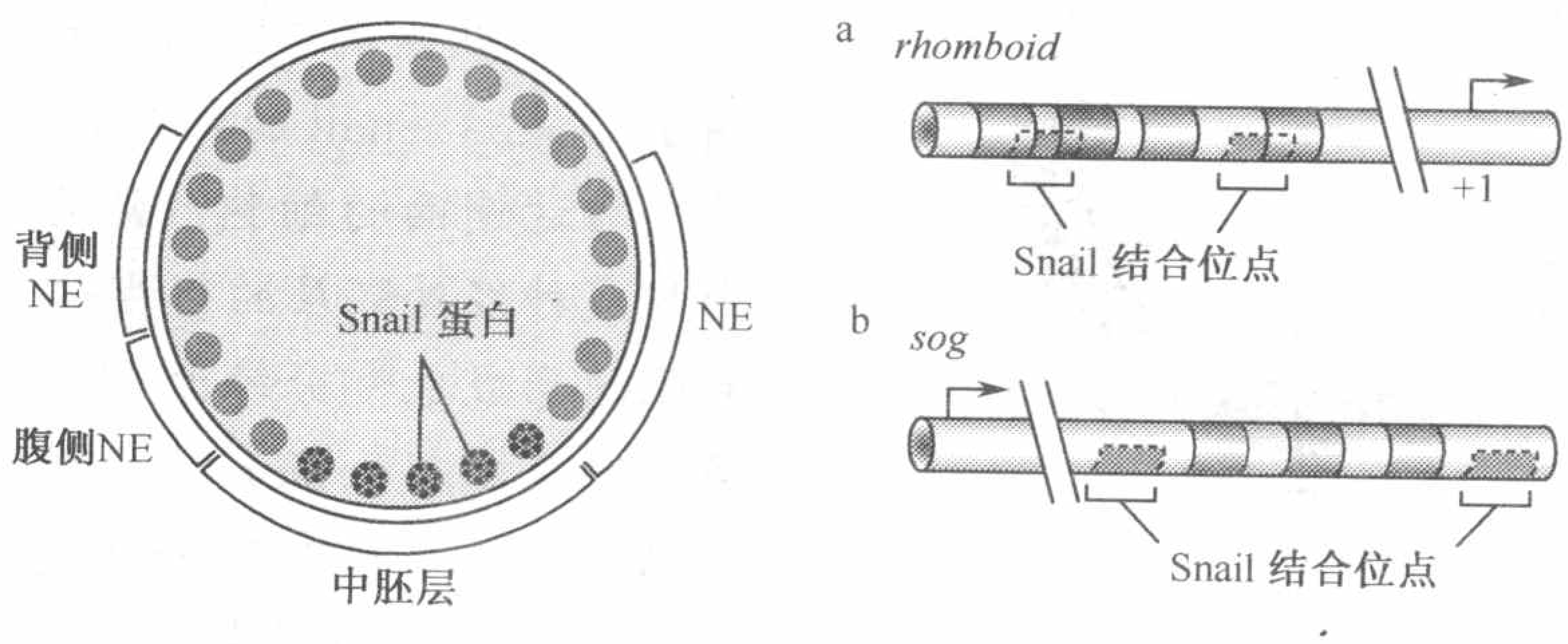


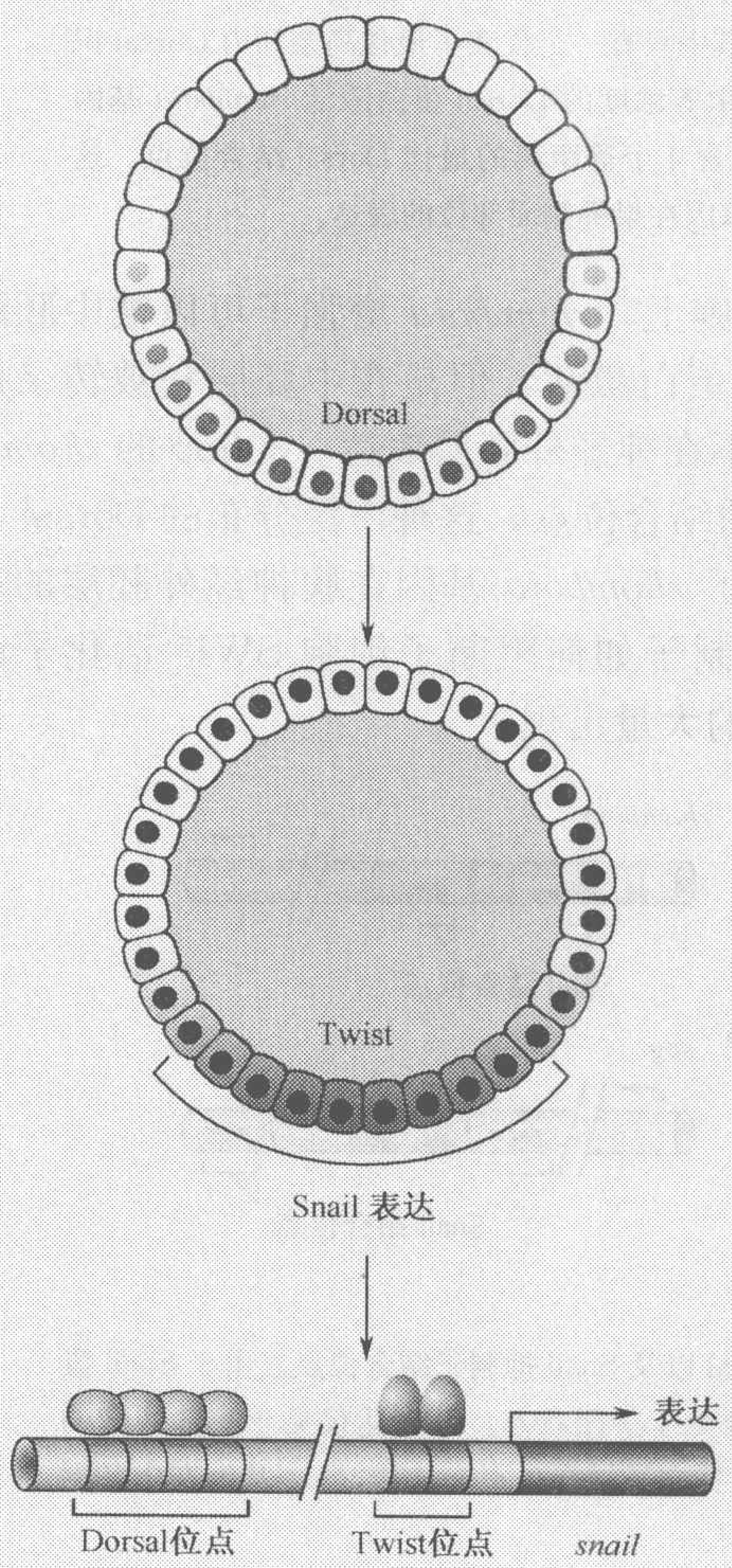
图 19-15　调节 DNA。 a. *rhomboid* 增强子包含 Dorsal 以及 Snail 抑制子结合位点。由于 Snail 蛋白只存在于腹部区域（中胚层），*rhomboid* 在中胚层处于关闭状态，其表达仅见于腹部区域的神经外胚层。b. 内含子区的 *sog* 增强子也包含 Snail 抑制子结合位点。这样，*sog* 基因在中胚层是关闭的，表达只见于侧面一片的腹、背神经外胚层。

早期胚胎侧面的腹背部神经外胚层 Dorsal 蛋白的低浓度足以活化 *sog* 基因。*sog* 基因的增强子是一段 400bp 的序列，位于第一内含子内（图 19-15b）。这个增强子含有

一排 4 个间隔的高亲和力的 Dorsal 结合位点，即使极低浓度的 Dorsal 蛋白也能使其占位。对 *rhomboid* 来说，Snail 抑制子的存在排除了 *sog* 基因在中胚层的表达，尽管那里 Dorsal 蛋白浓度最高。因此，不同阈值的 Dorsal 浓度梯度对基因表达的差异调控还要依赖于 Snail 抑制子和 Dorsal 结合位点亲和力的综合效应。

Dorsal 结合位点的占位取决于位点结合的亲和力，同时还依赖于 Dorsal 和其他增强子调节蛋白的相互作用。以上例子中，300bp 的 *rhomboid* 增强子得到腹神经外胚层中等浓度 Dorsal 蛋白的激活，它包含低亲和力的结合位点。由于与另一个活化子 Twist 的相互作用，中等浓度的 Dorsal 就足以结合这些位点。Dorsal 和 Twist 结合在 *rhomboid* 增强子的邻近位置。这两个蛋白质不仅在结合增强子时相互促进，一旦结合后还可以协同促进转录（框 19-5）。

框 19-5 发育中活化子的协同作用



框 19-5 图 1 Dorsal-Twist 协同作用模式。 Dorsal 显著的核浓度区间激活腹部 *twist* 基因表达，Dorsal 和 Twist 蛋白协同作用激活腹部和腹侧部位各种基因的表达。据推测，Dorsal 的作用是启动染色质修饰复合体同时 Twist 与 Mediator 或 TFIID 复合体反应促进转录。（来源：Stathopoulos A. and Levine M. 2002. Dorsal gradient networks in the *Drosophila* embryo. Dev. Biology 246：57-67，Fig. 2，p. 59. Copyright © 2002 with permission form Elsevier.）

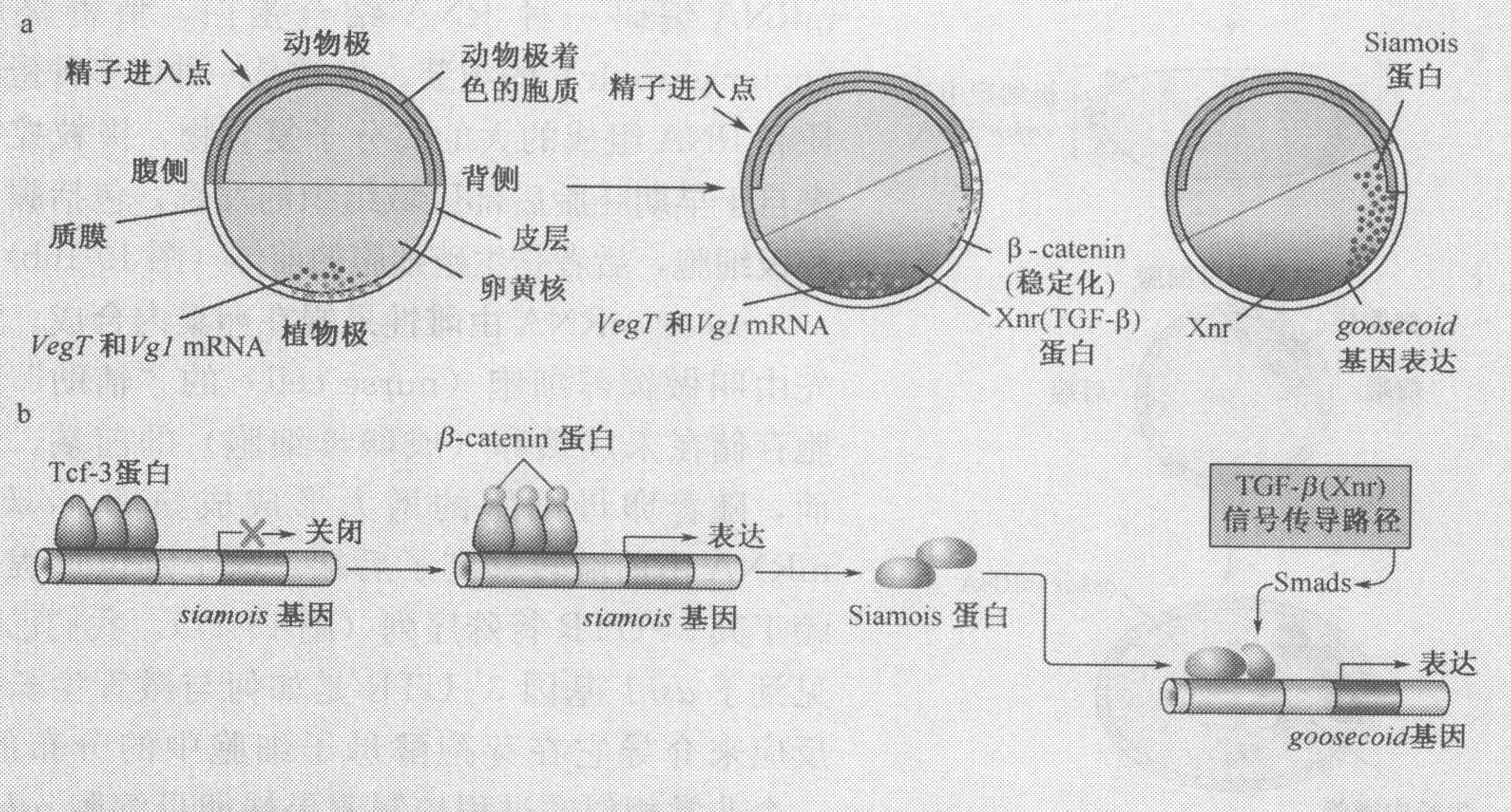
Dorsal 蛋白浓度仅仅 2 倍的差别就可以决定胚胎原细胞形成肌肉细胞还是形成神经元。这种细胞特征的调节转换受到 Snail 表达明显的边缘效应的限制（框 19-5 图 1）。表达 Snail 的细胞内陷形成中胚层，而更靠近侧面的细胞（缺乏 Snail 的表达）则形成神经外胚层的衍生体。明显的 Snail 界限的形成也部分缘于 Dorsal 和 Twist 的多重浓度梯度。Dorsal 的较宽的浓度梯度启动了 Twist 稍微窄一些的浓度梯度，然后二者协同作用限制 *snail* 的 5′端的调节 DNA 的活性表达（框 19-5 图 1）。

在 *snail* 基因转录起始点上游 1kb 处存在着 Dorsal 结合群和两个靠近 *snail* 启动子的 Twist 结合位点。由于 Dorsal 和 Twist 结合位点之间存在相当距离，二者不可能在 DNA 结合上有协同作用，但是有可能分别通过不同的转录限速复合体发挥作

用（混合协同，promiscuous synergy，第 17 章）。例如，Dorsal 可以通过募集如 SWI/SNF 等酶复合体改变染色体构象而打开 *snail* 基因 5′端的调节区。这样，有助于 Twist 的结合并进而将 TFIID-PolⅡ复合体募集到启动子核心区域（第 17 章）。本章随后还会讲到 Bicoid 和 Hunchback 怎样协同作用去活化 eve stripe 2。下面，我们会讲到脊椎动物运用相同的原则指导背（部）中胚层的形成。

青蛙卵背中胚层是原肠胚形成过程中控制中枢神经系统发育的重要信号分子的来源。背中胚层的形成依赖于未受精卵中 mRNA 如 *VegT* 等的定位。*VegT* 基因编码一种序列特异的转录因子能够在设定的中胚层区激活 *Xnr* 基因。*Xnr* 基因编码 TGF-β 信号分子，它在活化中胚层基因表达中发挥必要作用，但不足以单独起作用，这个作用是依赖 Xnr 和 Wnt 信号完成的。

受精后，出现一种皮层转换（cortical rotation）的过程，卵细胞细胞质向细胞膜方向转换（框 19-5 图 2a）。皮层转换使早期胚胎一侧的 β-catenin（β-连环蛋白）得到稳定，并继续形成对应于未来的背部。β-catenin 是细胞表面蛋白，通常在胞外分泌信号蛋白 Wnts 活化 Frizzled 受体后释放进入细胞核。然而，皮层转换可以替代 Wnts 的作用直接活化 Frizzled 受体释放 β-catenin。进入细胞核后，β-catenin 与序列特



框 19-5 图 2　青蛙胚背中胚层的特异化。 a. 青蛙卵包含一系列的定位的 mRNA，如 *VegT* 和 *Vg1*。*VegT* 编码 T 框 DNA 结合蛋白而 *Vg1* 编码活化素/TGF-β 信号分子（activin/TGF-β），二者促使 *Xnr* 在腹区的表达。受精后出现皮层转换导致 β-catenin 在未来的背区的固定化，Xnr 和 β-catenin 作用的界面界定了未来的背中胚层同时还引导一系列基因如 *goosecoid* 等的表达。b. 背部的 β-catenin 引发 *siamois* 基因的活化，后者编码一种同源框调节蛋白。Xnr 引发另一类调节蛋白 Smad 的活化。Smad 和 Siamois 只在背中胚层分布，二者协同作用激活 *goosecoil* 基因。[来源：a. Adapted from Alberts B. et al. 2002. Molecular biology of the cell，4th edition，p. 1211，f21-66. Copyright © 2002. Reproduced by permission of Routledge/Taylor & Francis Books，Inc. b. Adapted from Gilbert S. E. 2000. Development biology，6th edition，p. 322，Fig. 1025. Copyright © 2000 Sinauer Associates. Used with permission. And from Moon R. and Kimelman D. 1998. From cortical rotation to organizer gene expression. BioEssays 20:542，Fig. 3. Copyright © 1998. Used by permission of John Wiley & Sons，Inc.]

异的转录因子 Tcf 或 Pangolin 作用。

Tcf/β-catenin 复合物活化叫做 *siamois* 的靶基因，后者为同源异构调节蛋白。Siamois 的表达在整个背部区域都受到高浓度 β-连环蛋白（β-catenin）的调节，这种 Siamois 表达特征与中胚层的 Xnr 信号传导重叠（框 19-5 图 2b），重叠区间即背中胚层。Siamois 与 Xnr 协同作用活化背中胚层的靶基因，首批被活化的基因包括一个叫做 *goosecoil* 的同源异构调节蛋白。5′调节 DNA 包含 Siamois 以及 Smad 蛋白结合位点。Smad 是一类受 TGF-β 细胞表面受体活化而诱导的转录因子（框 19-5 图 2b）。在不进行信号传导时，Smad 与 TGF-β 受体的胞内部分结合处于非活化状态。信号传导时被释放进入细胞核。其结果，Smad 与 *goosecoil* 的 5′调节 DNA 结合。Smad 与 Siamois 在背中胚层协同作用激活 *goosecoil* 基因表达，表达部位表明在那里这两种活化子同时具有高浓度。

未受精卵前后极的定位 mRNA 启动分节

受精时，果蝇卵包含两种定位的 mRNA，一种为 *bicoid* mRNA，分布在前极；另一种为 *oskar* mRNA，分布在后极（图 19-16a）。*oskar* mRNA 编码一种 RNA 结合蛋白，负责极粒（polar granule）的组装。极粒是一种以多种蛋白质和 RNA 组成的大的高分子复合物，极粒控制来自于早期胚胎后部区域组织的发育，包括腹部和极细胞，后者是生殖细胞的前体（图 19-16b）。

oskar mRNA 由雌性果蝇在卵巢内合成。它先由叫做饲养细胞（nurse cell）的“辅助”细胞存储在未成熟卵（或卵母细胞）的前端。然而，随着卵母细胞的增大形成成熟卵，*oskar* mRNA 由前端转移到了后端。这个定位过程依赖于其 3′-UTR 特殊序列（图 19-17）。我们已经见到了 *ash1* 基因 3′-UTR 是如何与微管生长端反应来介导它在芽孢酵母子细胞中的分布的。一个非常相似的过程控制着果蝇卵母细胞 *oskar* mRNA 的分布。

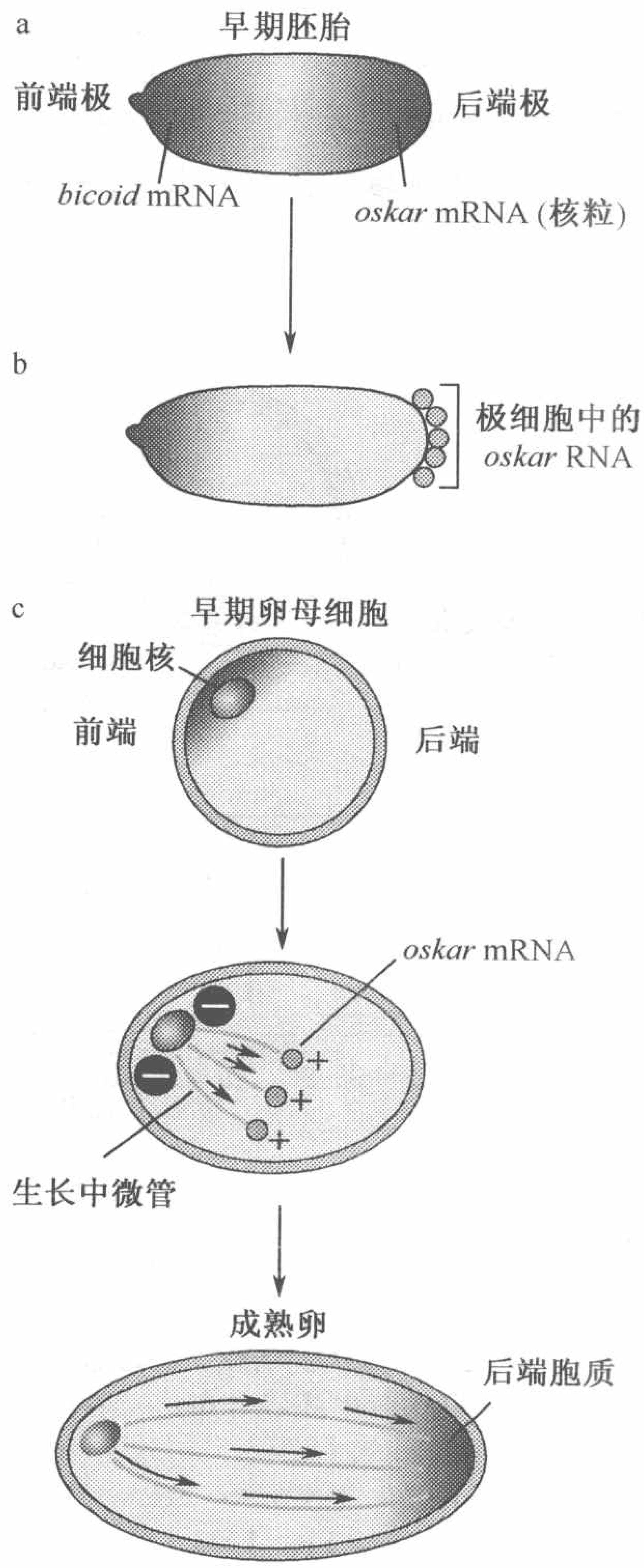


图 19-16 母系 mRNA 在果蝇卵和胚胎的定位。 a. 果蝇未受精卵包含两种定位的 mRNA，*bicoid* 在前端，*oskar* 在后端。b. Oskar 蛋白帮助协调在后端胞质的极粒的组装，细胞核在进入后端后以出芽的方式将胚胎后端转化为极细胞。c. 在果蝇卵形成时，极性微管从卵母细胞的核延展向后端的细胞质，*oskar* mRNA与适配体蛋白结合进而与微管反应将其转运入后部胞质。这里的“－”和“＋”符号标明微管的生长方向。

果蝇卵母细胞具有高度极性，其细胞核分布在前端；生长的微管从核伸展到胞质的后端。*oskar* mRNA 与微管生长的正端的适配体蛋白反应转运出细胞核存在的前端到达后部的细胞质。受精后，得到定位 *oskar* mRNA 的细胞（同时包括极粒）形成极细胞。

bicoid mRNA 的在非受精卵前端的定位也依赖于其 3′-UTR 的特殊序列。*oskar* mRNA 和 *bicoid* mRNA 的 3′-UTR 的序列并不相同，因而它们分别与不同的适配体蛋白反应定位于卵的不同部位，3′-UTR 在 mRNA 定位中的重要性在以下实验中得到阐明。如果 *oskar* mRNA 的 3′-UTR 被 *bicoid* mRNA 的取代，那么杂合的 *oskar* mRNA 就会留在前端（与 *bicoid* mRNA 正常行为一样）。这种错误定位足以诱导早期卵细胞在异常的位置形成极细胞（图 19-17）。此外，错位的极粒还会抑制头部组织分化需要的基因的表达。因而，应该形成头部组织的细胞转化成了生殖细胞。

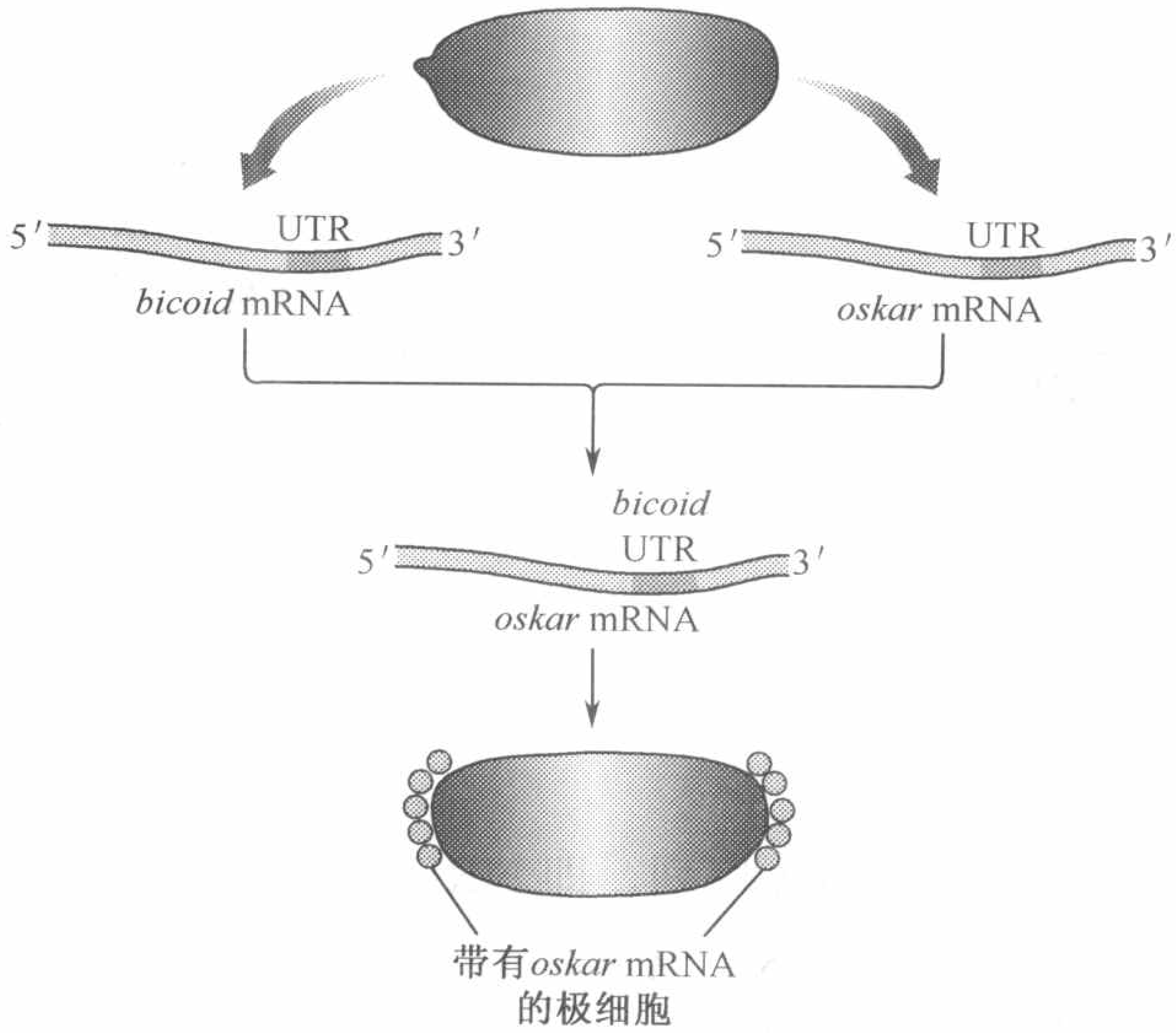


图 19-17　*oskar* mRNA 和 *bicoid* mRNA 包含不同的 3′-UTR 序列。 *bicoid* mRNA 的 3′-UTR 序列导致它在前端的分布，而 *oskar* mRNA 的 3′-UTR 序列导致它在后端的分布。改造后的 *oskar* mRNA包含 *bicoid* mRNA 的 3′-UTR 将向正常 *bicoid* mRNA 一样定位于前端。*oskar* mRNA 这种错位造成在前端形成极细胞。此外，错位的极粒还会抑制头部组织分化需要的基因的表达，因而，应该形成头部组织的细胞转化成了生殖细胞，定位于后部胞质的正常的 *oskar* mRNA 也会促使后端形成极细胞。

Bicoid 和 Nanos 调节 hunchback 基因

Bicoid 调控蛋白的合成完成于细胞化之前。因此，由前极合成的部位开始 Bicoid 扩散开来，并沿着早期胚胎的长度分布形成了一个宽泛的浓度梯度。高、中浓度的 Bicoid 均足以激活 hunchback，而 hunchback 对胚胎的精细分区起至关重要的作用（图 19-18）。*hunchback* 基因实际上是由两个启动子控制转录的：一个由 Bicoid 梯度激活，另

一个在卵母细胞的发育过程中控制基因的表达。后者，也就是“母系”启动子引导合成的 hunchback mRNA 均匀地分布在未受精的卵子细胞质内。这种母性的转录子的翻译在后区被一个叫做 Nanos 的 RNA 结合蛋白所阻止（图 19-18）。Nanos 蛋白只在后区被发现，因为其 mRNA 的 3'UTR 和极性颗粒之间的相互作用，所以选择性的存在于这个区域。

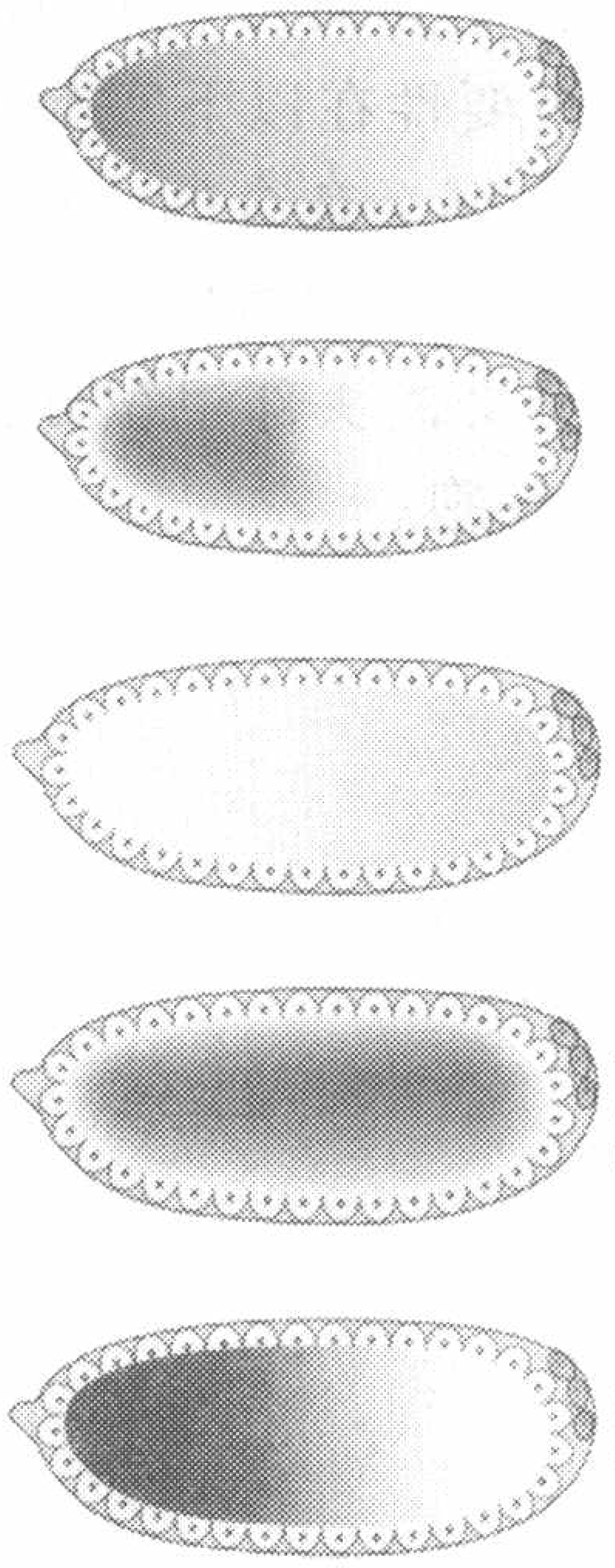


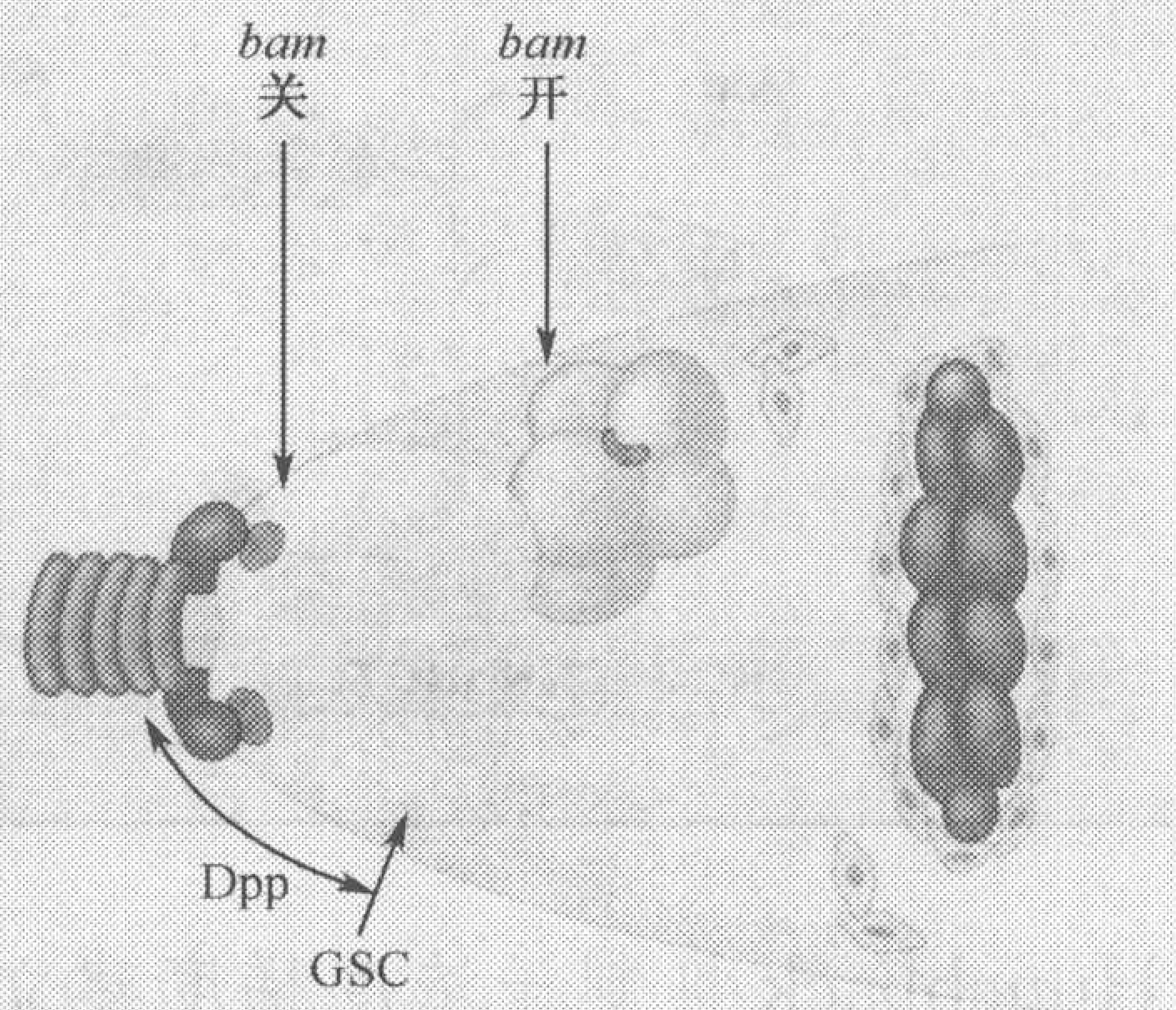
图 19-18 Hunchback 蛋白质的浓度梯度及 Nanos 对翻译的抑制。由于高、中浓度的 Bicoid 梯度能够激活 hunchback，宽泛的存在于前后期的 Bicoid 蛋白质浓度梯度使 hunchback 基因的表达产生了一个明显的阈值。Nanos mRNA 与极性颗粒联系在一起，翻译后在后区扩散并形成了一个浓度梯度。母性 hunchback mRNA 遍布于早期的胚胎中，但其翻译却被结合在 hunchback 3'UTR 特异序列的 Nanos 蛋白捕获。因此 Nanos 蛋白质的浓度梯度导致了前区相应的 hunchback 浓度梯度。

Nanos 蛋白质结合特异的 RNA 序列，即 NREs（Nanos 反应元件），NREs 位于母性 hunchback mRNAs 的 3′UTR 区，这种结合导致了 hunchback 的减少。Poly-A 尾巴，反过来破坏 RNA 的稳定并抑制其表达（第 14 章）。因此，我们可以看到 Bicoid 浓度梯度在胚前半期激活了 hunchback 启动子，而 Nanos 却在后期抑制了母性 hunchback mRNA 的翻译（图 19-18）。这种 hunchback 表达的双重调控产生了一个 hunchback 蛋白浓度的急剧升降，其中胚胎的前半期为浓度最高，而在后半期急剧地下降。

框 19-6 干细胞

干细胞通常是“全能的”，也就是它们有能力形成各种不同类型的细胞。事实上，哺乳动物的胚胎干细胞，内细胞团（ICM）中的细胞，可产生成熟个体中所有复杂的组织，包括不同的血液细胞及构成皮肤、肌肉和内脏器官的不同类型的细胞。由于巨大的在疾病治疗中的应用潜力，干细胞引起了广泛的兴趣和关注（框 17-6）。也许某一天，干细胞有可能用来替换我们的“磨损部件”，比如胰岛素用于胰腺恶化的糖尿病患者。干细胞拥有无限自我更新能力。然而，在疾病治疗中，为了降低可能的癌症发生的机率，重要的是限制干细胞的增殖。

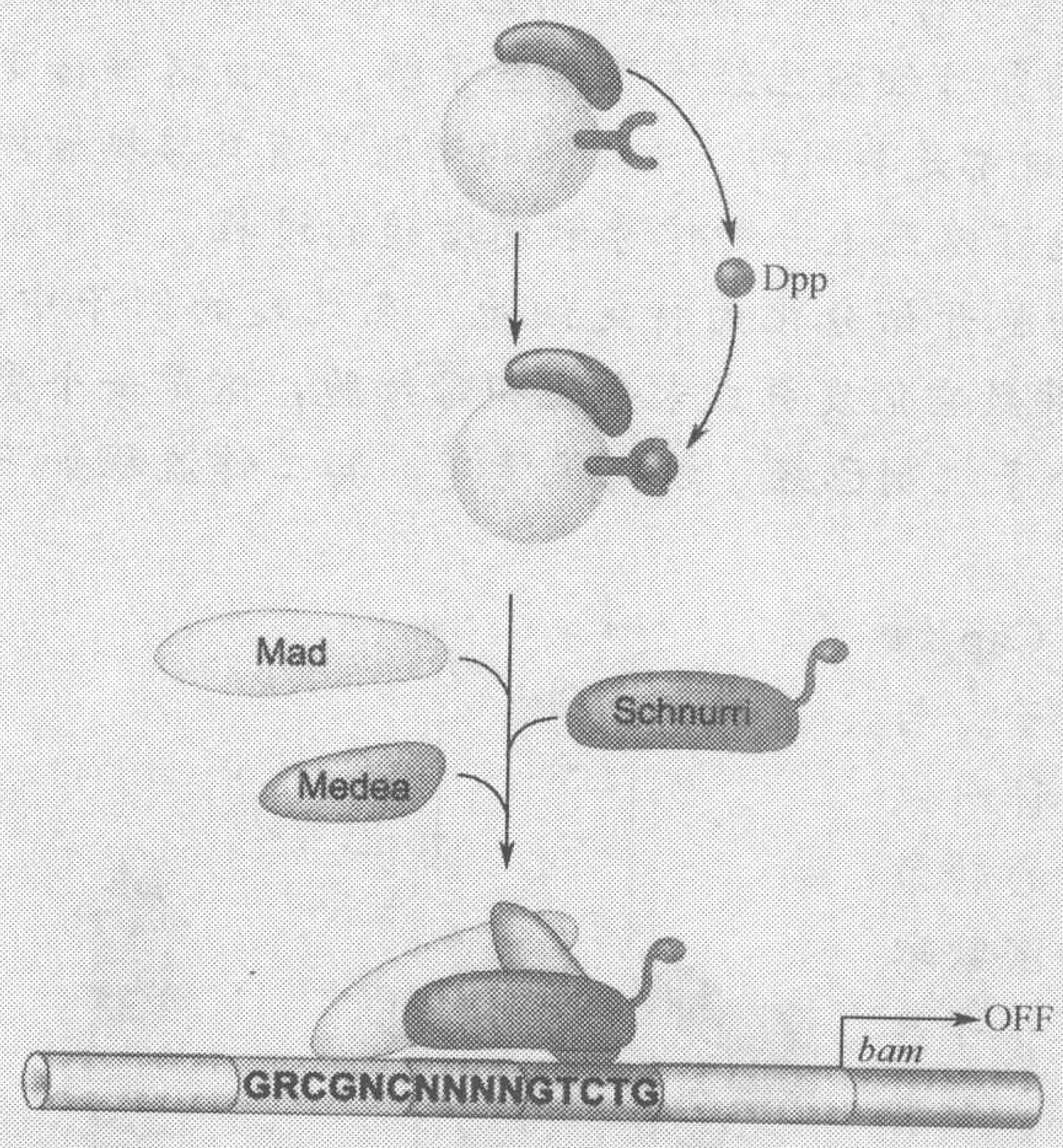
在果蝇中，卵细胞或者卵母细胞起源于一种叫做生殖系干细胞（GSC）的干细胞前体。关于由GSC如何转变成果蝇卵巢内的卵母细胞已经有了相当了解，而且这种转变机制的许多方面对果蝇和人类的干细胞来说都是共同的。干细胞的增殖只有直接接触被称之为“niche”的特定的细胞，这类细胞可以产生一种信号触发增殖的过程。当干细胞从niche上解离下来后，增殖停止，细胞开始分化为特定细胞。在果蝇的例子中，GSCs从卵巢中肺泡细胞上解离下来后，导致它们发育成未分化卵母细胞，这是一个受信号介导的阻遏调节过程。这一过程的分子机制已经了解得很清楚，其工作流程如下所述。



框19-6图1 卵母细胞发育中bam的表达。该示意图描绘了bam mRNA及蛋白质表达方式及其分布。Cap细胞（用紫色表示）分泌Dpp，激活其生殖层干细胞（黄色表示GSCs）上的受体，从而引发信号传递过程最终抑制了bam的表达。随着GSGs从Cap细胞上解离，Dpp信号传递的中断，bam mRNA的表达导致了bam蛋白质在细胞质中的高表达。在Bam蛋白质的存在时，分离的子细胞发育成卵母细胞前体（绿色）并进一步分成8细胞囊胞（深绿色）。（经许可改编自Chen D. 和Mckearin D. M. Development 130：1159-1170，Fig. 1.© Company of Biologists.）

位于果蝇卵巢内的肺泡细胞，也叫Cap细胞，分泌一种可被传播的Dpp信号分子。与GSCs结合后Dpp被激活从而导致了一种叫做bam的重要的调控基因的沉默：当bam的转录被阻断后，GSCs开始增殖。bam表达的沉默取决于Cap细胞与GSCs的直接接触。这类似于昆虫神经系统里Notch信号传递的激活过程。随着GSCs的增殖，一些子细胞开始从Cap细胞上解离下来，从而不再是Dpp信号分子的靶目标。在没有信号的情况一下，bam转录被激活，细胞停止增殖，转而分化成为卵母细胞（框19-6图1）。

干细胞的增殖及卵母细胞的分化之间的选择取决于bam表达这一开关机制。现在已经知道，这种调控是由bam 5′调控区域的一个沉默元件介导的，即GRCGNC（N）5GT CTG序列（框19-6图2）。Dpp信号起始了两个Smad调控蛋白Mad和Medea的核转运。这些蛋白结合到沉默元件的2个半位点，接着加上一个转录阻遏蛋白，ZF6-6也称Schnurri，阻止bam的转录。但这个转录阻遏蛋白及随后发生的抑制bam转录仅仅在GSCs与Cap细胞保持接触的情况下才会发生。结果使得这些细胞得以分化并产生更多的干细胞。相反地，当GSC子细胞从Cap细胞上解离下来，barn的转录被激活，这是因为导致基因沉默信号通路被打断了。在这些细胞内，Dpp受体处于失活状态（因为信号传递被阻断）所以不会结合到5′沉默元件上或者加上Schnurri受体。在这种条件下，bam得到了表达，子细胞不再增殖而是分化成卵母细胞。要求肺泡细胞和干细胞之间的直接接触以及所导致的信号介导的阻遏过程可能是保证干细胞持续分化的共同机制。



框 19-6 图 2 Dpp 途径有效的抑制关键的发育基因在 GSC（黄色）中。 由 Gap 细胞分泌的 Dpp 与 Dpp 受体的结合发出了信号，促进 Mad（黄色）和 Medea（蓝色）转运到核中。受到抑制的靶基因（此处以 Bam 为例）含有顺式作用沉默元件（SE），它们与 Mad、Madea 结合，共同招募 Schnurri（用粉色表示）从而有效地阻止转录。（经允许引自 Pyrowolakis G. et al. 2004 Dev. Cell 7：299-240，图 7 Elsevier 版权所有）

Hunchback 抑制子浓度梯度建立了 Gap 基因表达的不同限度

Hunchback 作为转录抑制子建立了所谓“间隔”基因，如 *krüppel*、*knirps*、*giant* 的不同表达的界限（在框 19-4 已有讨论）。我们还会看到 Hunchback 与这些间隔基因蛋白一起作用产生基因表达的分节条，即开始将胚胎分成重复的分节体节过程的第一步。

Hunchback 蛋白沿未来的胸部向腹部形成一个深的浓度梯度，高浓度的 Hunchback 蛋白抑制 *krüppel* 基因的表达，低浓度时抑制 *knirps* 和 *giant* 基因的表达（图 19-19a）。我们见过 Bicoid 和 Dorsal 活化子的不同亲和力产生不同的基因表达的阈值。Hunchback 抑制子的作用方式颇为不同，Hunchback 抑制子的数目成为了形成 *krüppel*、*knirps*、*giant* 独特基因表达类型的主要决定因素（图 19-19b）。*krüppel* 增强子只有 3 个 Hunchback 结合位点受到高浓度的 Hunchback 抑制。相反，*giant* 增强子有 7 个 Hunchback 结合位点受到低浓度的 Hunchback 抑制，其中的作用机制还未被揭示，或许抑制的不同阈值产生于单个 Hunchback 抑制域的叠加效应。

Hunchback 和间隔蛋白产生了基因表达的分节条

从 *bicoid* 和 *oskar* 定位开始的调节级联反应在 *even-skipped*（或简称 *eve*）的“成对”（pair-rule）基因表达时达到顶点。*eve* 基因以一种 7 轮回条或成对的方式在胚胎表达（图 19-20）。每个 *eve* 条包含 4 个细胞，同时条间隔也有 4 个细胞，但不表达 *eve*，这些条预示着未来体节形成的位置。

Eve 蛋白的编码序列并不长，只有约 2kb 个碱基对，而其两边的调节区却有 12kb，其中 4kb 在 5′端转录起始点上游，8kb 在 *eve* 基因 3′端外围（图 19-20）。5′端调节区启

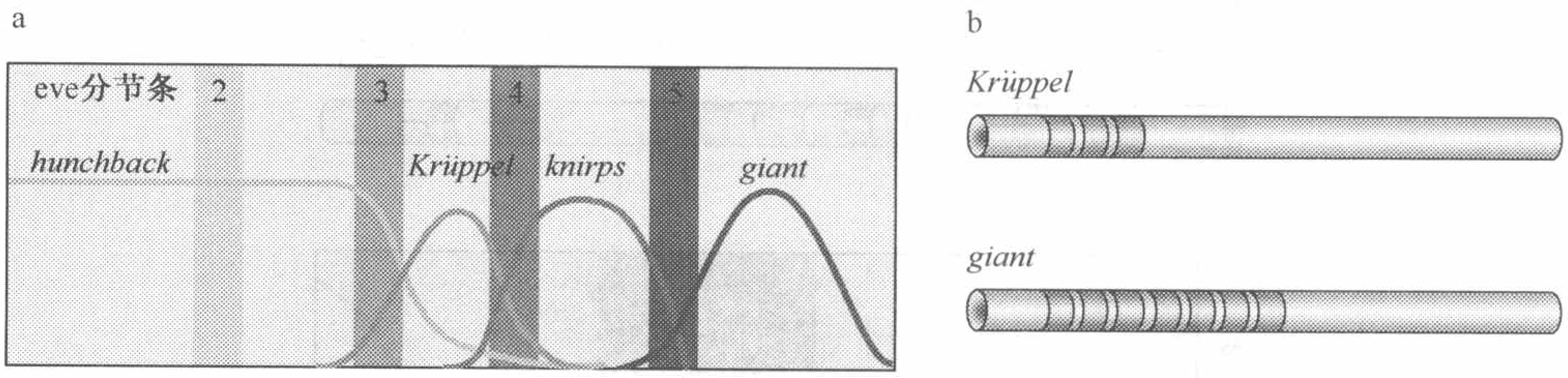


图 19-19　Hunchback 生成连续的间隔的基因表达类型。 a. 前后部位 Hunchback 抑制子形成的浓度梯度建立了 *krüppel*、*knirps*、*giant* 基因表达的不同的界限。高浓度抑制 *krüppel* 而低浓度抑制 *giant*。b. *krüppel* 和 *giant* 的 5′端调节 DNA 含有不同数目的 Hunchback 抑制子结合位点，*krüppel* 有 3 个位点而 giant 有 7 个位点。*giant* 增强子中增加的 Hunchback 抑制子结合位点对应于它受低浓度 Hunchback 抑制的效应。（来源：a. Redraw from Gilbert S. E. 1997. Developmental biology，5th edition，p. 565，Fig. 14-23. Copyright©，1997 Sinauer Associates. Used with permission.）

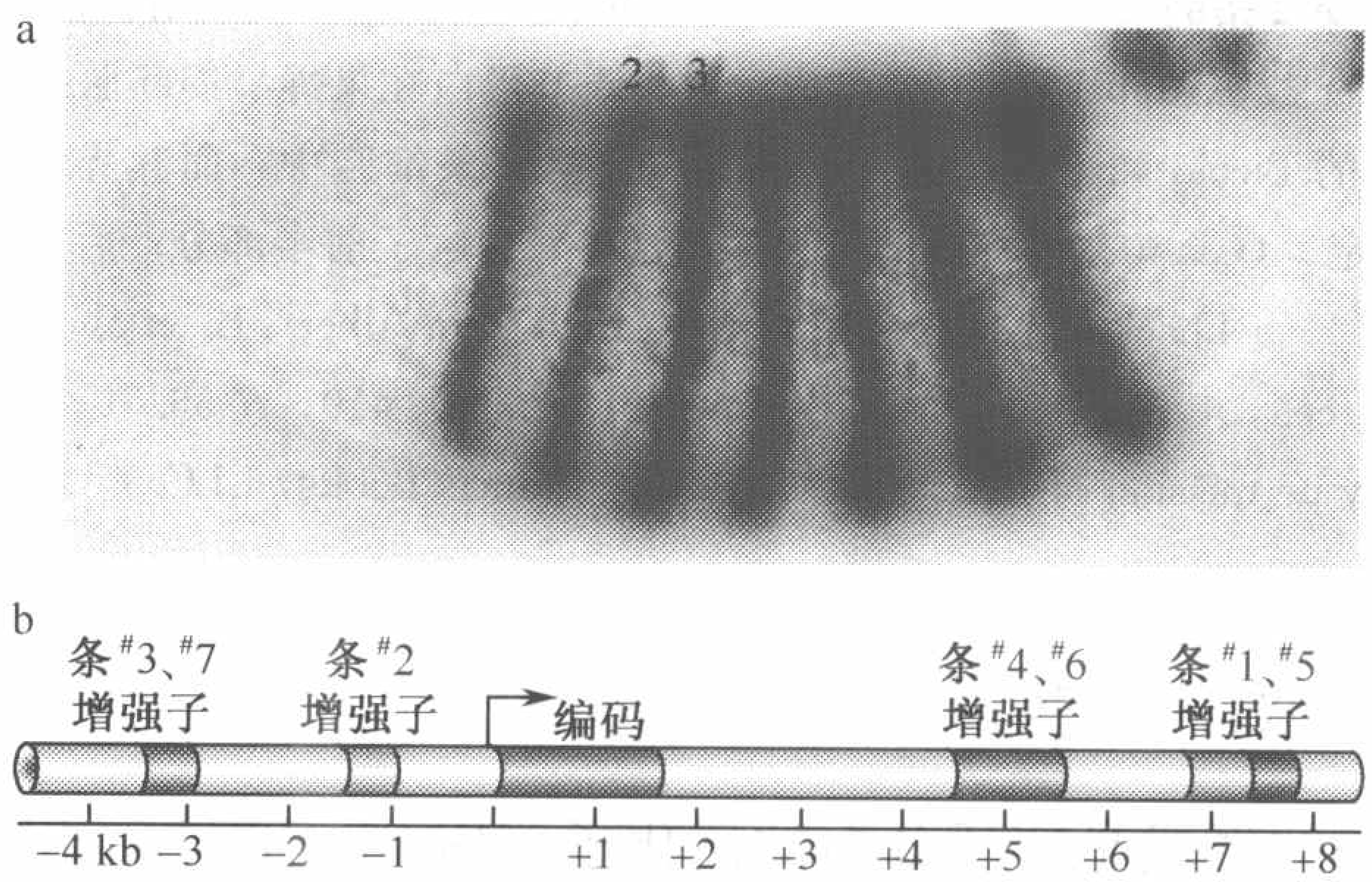


图 19-20　发育胚胎中 *eve* 基因的表达。 a. 早期胚胎 *eve* 基因表达类型。b. *eve* 位点包含 12kb 的调节 DNA 序列，其 5′端包含 2 个增强子，控制第 2、3、7 条的表达，每个增强子的长度为 500bp。3′端包含 3 个增强子，控制第 1、4、5、6 条的表达。这 5 个增强子形成了早期胚胎的 7 个不同的 *eve* 基因表达条。[来源：a. Image courtesy of Michael Levine.]

动第 2、3、7 条的表达，而 3′端调节第 1、4、5、6 条的表达。这 12kb 的区间包含 5 个增强子，其结果是形成早期胚胎的 7 个不同的基因表达条。每一个增强子只控制第 1～2 个条的表达，这里要讲到第 2 条表达的增强子调节。

第 2 条增强子有 500bp，分布在转录起始点上游 1kb 的位置，它包含 4 种调节蛋白，Bicoid、Hunchback、Giant 和 Krüppel 的结合位点（图 19-21）。我们已知 Hunchback 在调节间隔基因表达时起到抑制子的作用；而对第 2 条的调节则起到活化子的作用。至于 Hunchback 如何既做活化子又做抑制子，我们以后还会讲到。原则上，Bicoid 和 Hunchback 存在于胚胎的前半部分可以活化第 2 条的增强子，Giant 和 Krüppel 作为抑制子界定了第 2 条的类型即前后体的分界（图 19-21）。

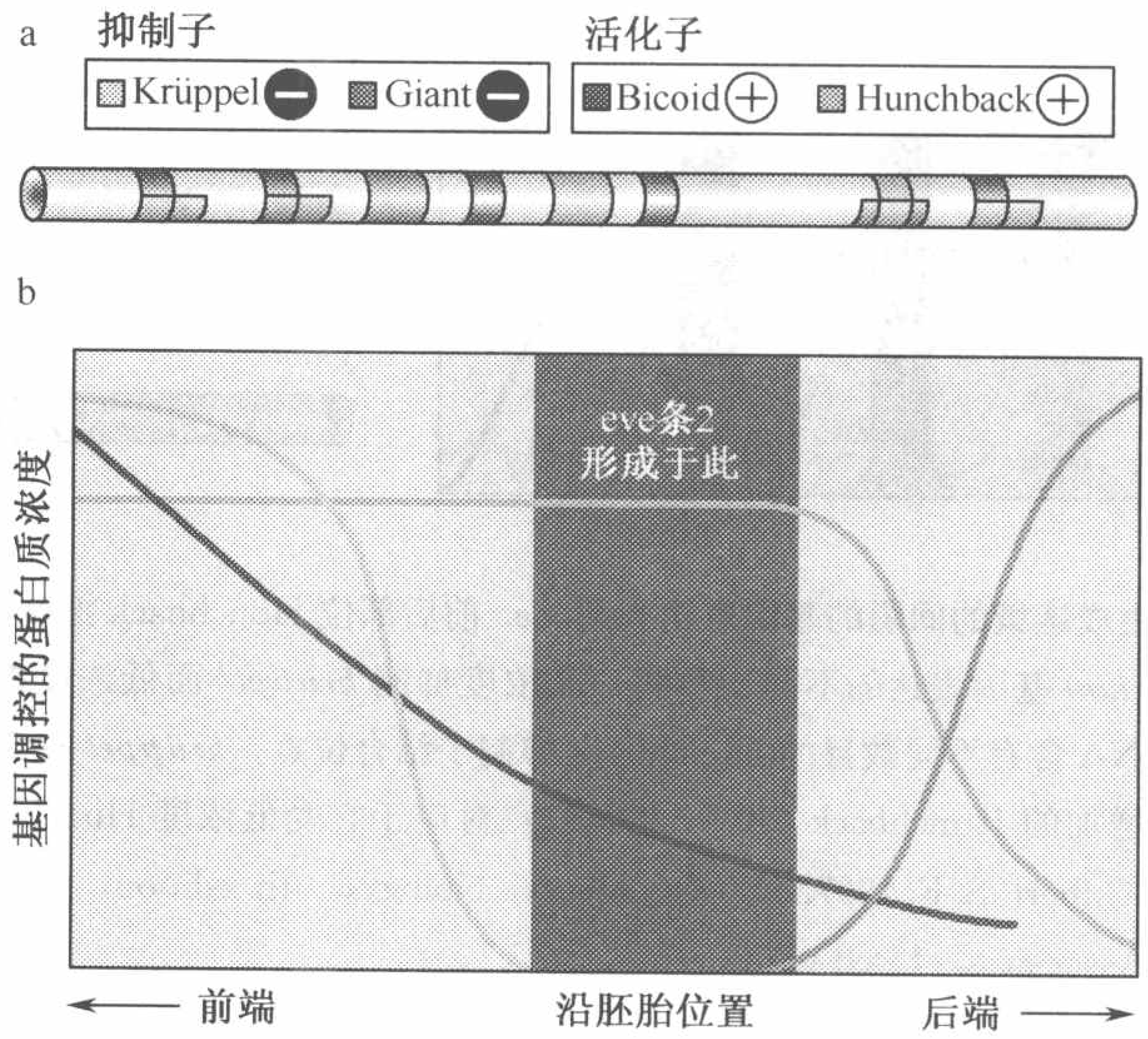


图 19-21 ***eve* 第 2 条的表达调节。** a. 500bp 的增强子包含总共 12 个结合位点，对应于 Bicoid、Hunchback、Krüppel 和 Giant 蛋白。这些调节蛋白在果蝇早期胚胎中的分布的总结见 b。高浓度的 Bicoid 和 Hunchback 可见于 *eve* 第 2 条细胞中，而其界面则由 Giant 和 Krüppel 来界定。(Giant 在前、后部位表达，这里只展示了前半部分；后半部分的类型未展示，它是由 Hunchback 调节的。)（来源：Adapted from Alberts B. et al. 2002. Molecular biology of the cell，4 th edition，a. p. 409，f7-55，b. p. 410，f7-56. Copyright © 2002. Reproduced by permission of Routledge/Taylor & Francis Books，LLC.）

间隔抑制子浓度梯度引起基因表达的条带化

Eve 第 2 表达条的形成是广泛分布的活化子（Bicoid 和 Hunchback）和局部分布的抑制子（Giant 和 Krüppel）相互作用的结果，同样的机制也适用于其他 *eve* 增强子的调节。例如，在早期胚胎中 *eve* 第 3 表达条的增强子受到普通的转录活化因子的活化。表达条的边界由局部分布的间隔抑制子界定：前界由 Hunchback 确定而后界由 Knirps 确定（图 19-22）。

框 19-7 动物发育和进化中的顺式调控序列

顺式调控序列以模块方式存在于动物基因组内。通常来说，一个复杂的表达模式中各组分有独立的增强子。考虑到一个发育中的老鼠胚胎内某一个基因在不同的组织和器官中表达，比如肝脏、胰腺和脑垂体中。有趣的是这个基因带有用于不同部位表达的不同的增强子。我们已经知道 *eve* 座位中包括 5 个独立的增强子位于 5′和 3′侧翼区（图 19-20）。每一个增强子决定着早期果蝇胚胎 7 种 eve stripes 中一种或两种模式的表达。正如我们下面将讨论的，这种模块化组织形式有利于顺式调控序列在进化中的形成多样化的形态。

模块化组织的控制基因的多效性

在进化过程中，基因表达的方式是如何改变的呢？越来越多的证据表明，关键性的激活蛋白结合位点处核苷酸的改变可以消除基因在特定组织或细胞进化中的表达。这里举个棘鱼腹鳍的例子。在自然界中，有无腹鳍的棘鱼变异体。当它们与有腹鳍的个体配对时，有可能鉴定其负责腹鳍减少的主要基因位点。该位点位于 *pitxl* 基因的 5′侧翼区域。Pitxl 是一个发育调控基因，在老鼠发育的几个不同的组织，包括胸腺，嗅窝和后肢中，是必不可少的。在棘鱼中，包含腹鳍（后肢）增强子的关键性的激活蛋白结合位点造成的点突变似乎导致了腹鳍的减少（框 19-7 图 1）。这些突变打断了腹鳍发育中的基因表达，但它们不干预在胸腺、嗅窝和其他组织中调控 *Pitxl* 基因的增强子的活性，在这些部位中，*Pitxl* 基因呈表达状态。

顺式调控区域模块中的特定改变也决定着不同种类果蝇的颜色形式的进化。传统的黄色位点（y）对着色至关重要，简单的基因突变就可造成黄色果蝇不产生局部的黑斑。由不同的增强子所调控的 y 参与鬃毛、翅膀和腹部的基因表达的方式也是如此。

成年黑腹果蝇（尤其是雄性）的后腹部片段含有强色素沉着。这种沉着是由 y 腹部增强子 Hox 蛋白质 Abd-B 的直接激活造成的。缺乏腹部片段的果蝇，如 kikkawai，在关键的 Abd-B 激活位点有一个点突变。由此导致了腹部 y 的表达缺失及所观察到的色素沉着缺失现象。

在翅膀中，有一个调控 y 表达的单独的增强子。在一些种类的果蝇中，这个增强子导致的成年雄性翅膀中含有一个特别的四分之一翅膀大小的色素沉着点（框 19-7 图 2）。这个点是求爱仪式的一个关键组成部分。而不含这个色素点的果蝇中翅膀的增强子中含有点突变，从而导致 *y* 基因的活性的部分丧失，但并不影响 *y* 基因在其他组织，如鬃毛、腹部表皮的活性。

阻遏蛋白位点的改变导致基因表达的显著变化

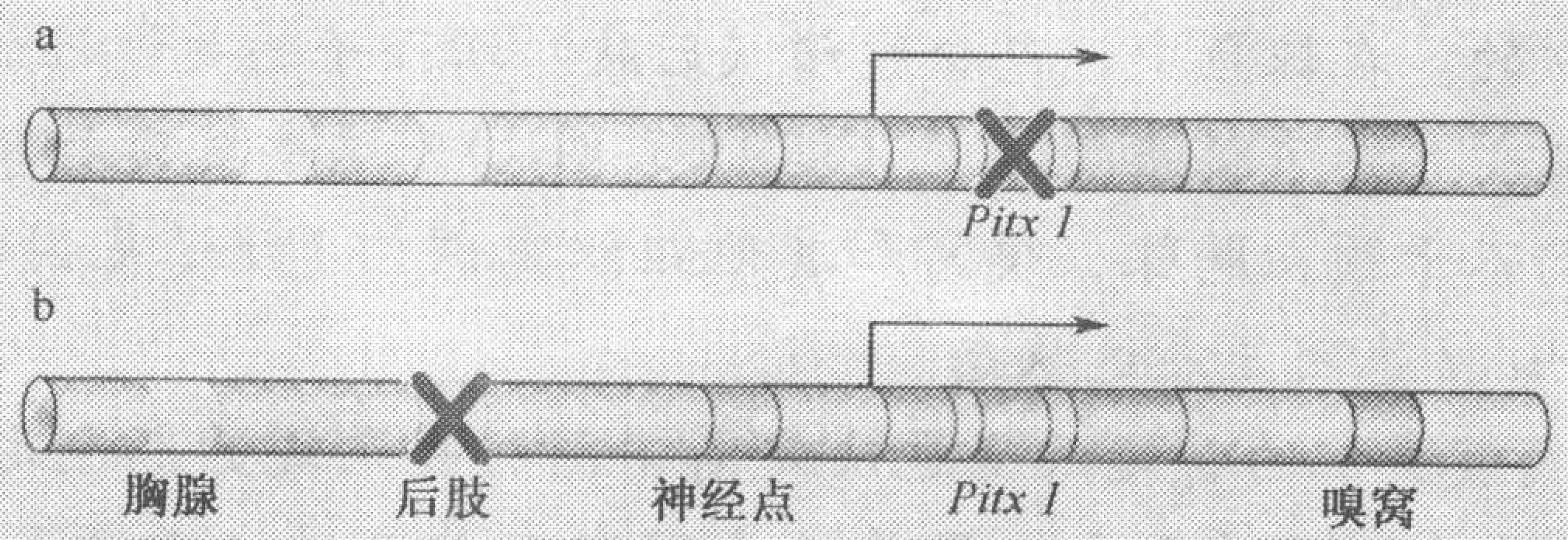
在不连续的增强子模块中，关键激活位点的丧失可以解释 Pitxl 局部表达的缺失及 y 基因的活性。新的基因表达模式的产生可能是阻遏元件的丢失所引起的。

在果蝇早期胚胎中，绝大多数或全部有活性的增强子都有阻遏蛋白结合位点，它们负责形成明显区分的基因表达边界。例如，eve stripe-2 增强子带有 Giantt 和 Kruppel 阻遏蛋白的结合位点，导致前后期基因表达的急剧的变化（图 19-21）。这些位点的突变使正常的表达模式发生急剧的扩展：形成了一个宽范围的表达，而不是局部条纹状的表达。

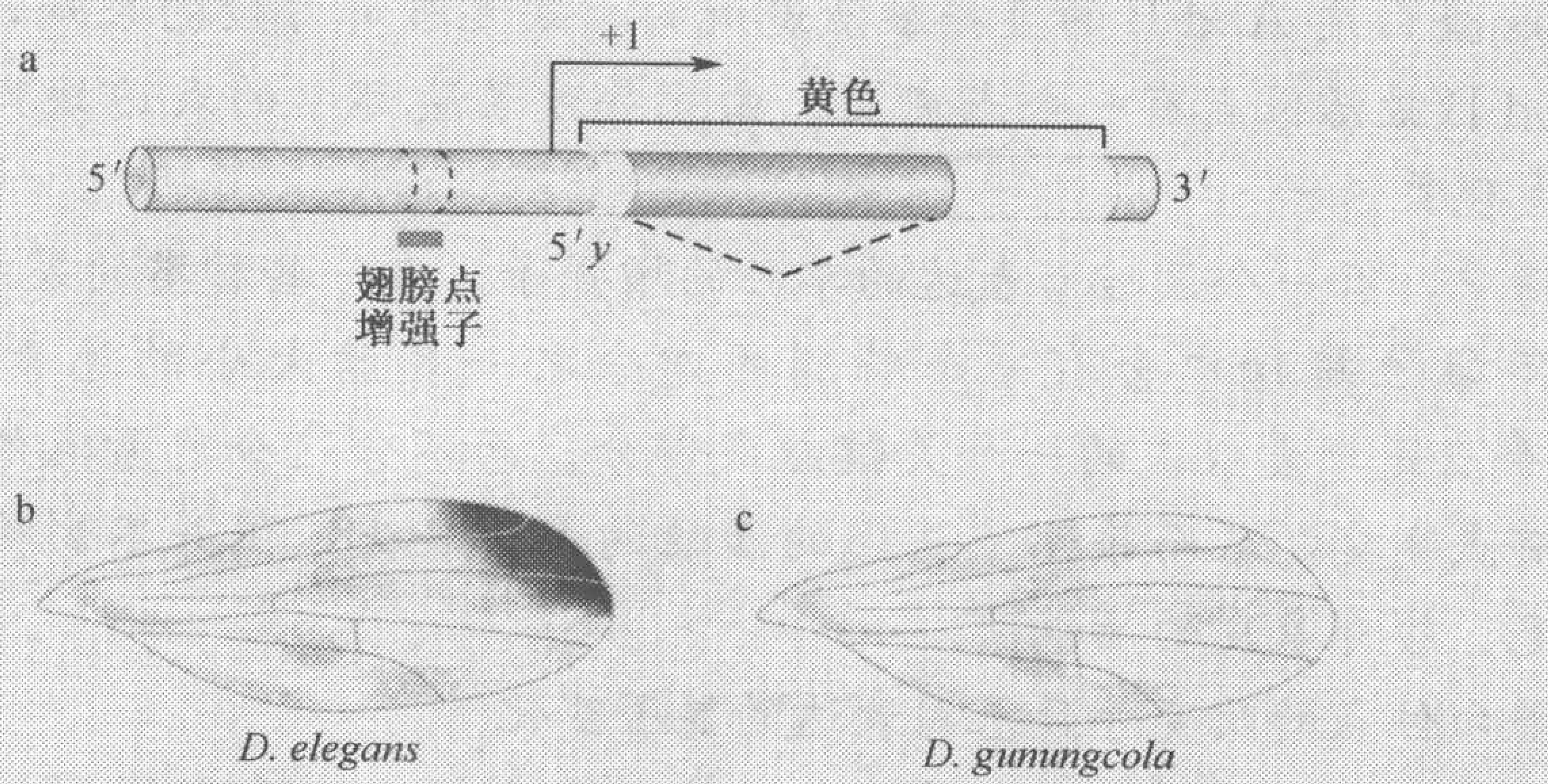
人类中的乳糖酶基因是一个可能的阻遏元件参与进化的范例。在大部分灵长类动物中，在母乳喂养期的婴儿小肠中的 LCT 基因高表达。但是，这个基因会在青春期后关闭。成年人中乳糖酶基因的持续表达是不常见的。牧区人群乳糖酶基因的持续表达与他们断奶后长期食用牛奶有关。某些特定的人群中具有较高的 LCT 基因表达是因为在位于

乳糖酶基因 LCT 基因 5′端的 MCM6 基因内元中有核苷酸的替换（框 19-7 图 3）。

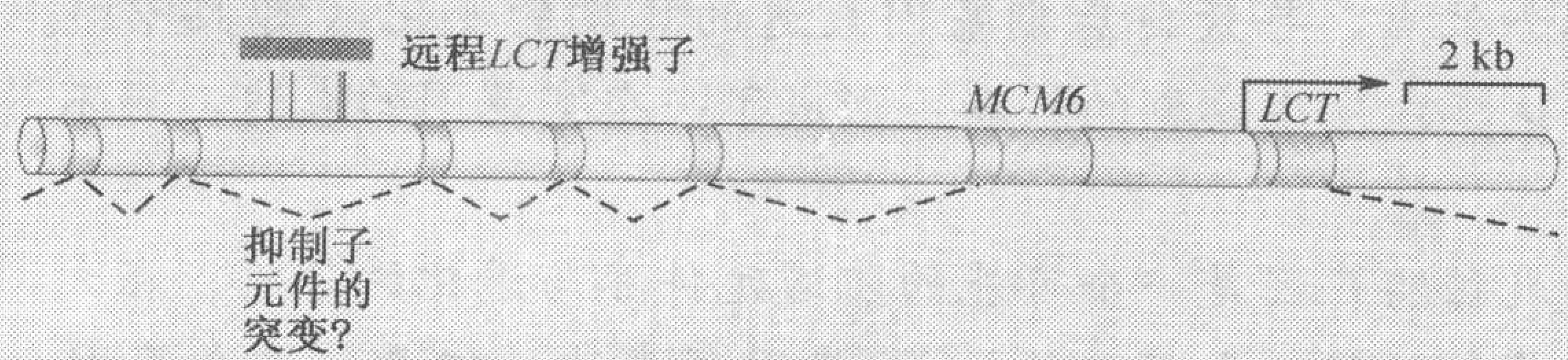
这些核苷酸的替换有可能破坏阻遏元件正常地结合在沉默蛋白上，从而使青少年和成人的小肠中抑制乳糖酶基因的表达遭到破坏。这种起关键顺式作用元件的缺失与棘鱼 *Pitx1* 基因的后肢/腹鳍增强子失活或果蝇 y 基因的腹部/翅膀增强子的失活类似。但是，在乳糖酶基因的例子中，因为阻遏元件的缺失使得乳糖酶活性得到了暂时的持续，实际是一种新的基因表达方式。



框 19-7 图 1 发育调控基因 *Pitx1*。(a) 此图显示了 *Pitx1* 基因以及 5′上游序列的结构。这里显示的是在基因编码区（第二个外显子）发生的（使实验室老鼠）致死的无义突变。(b) 在野生型棘鱼中，5′上游序列中的一个关键的调控突变导致了腹鳍的缩小。



框 19-7 图 2 果蝇的黄色基因位点 y。(a) 此图显示了黄色基因位点的结构以及上游调控序列（增强子序列）。(b) 一种成年雄性果蝇翅膀上的正常的色素沉着点（求偶点）。(c) 缺失色素沉着的一类果蝇的翅膀，该物种带有 5′增强子的突变位点。



框 19-7 图 3 乳糖酶基因的结构及其 5′上游调控区

eve 第 4 表达条增强子也受到 Hunchback 和 Knirps 的抑制，这些抑制子的不同浓度发挥着不同的作用。低浓度的 Hunchback 不足以抑制 *eve* 第 3 表达条增强子但足以

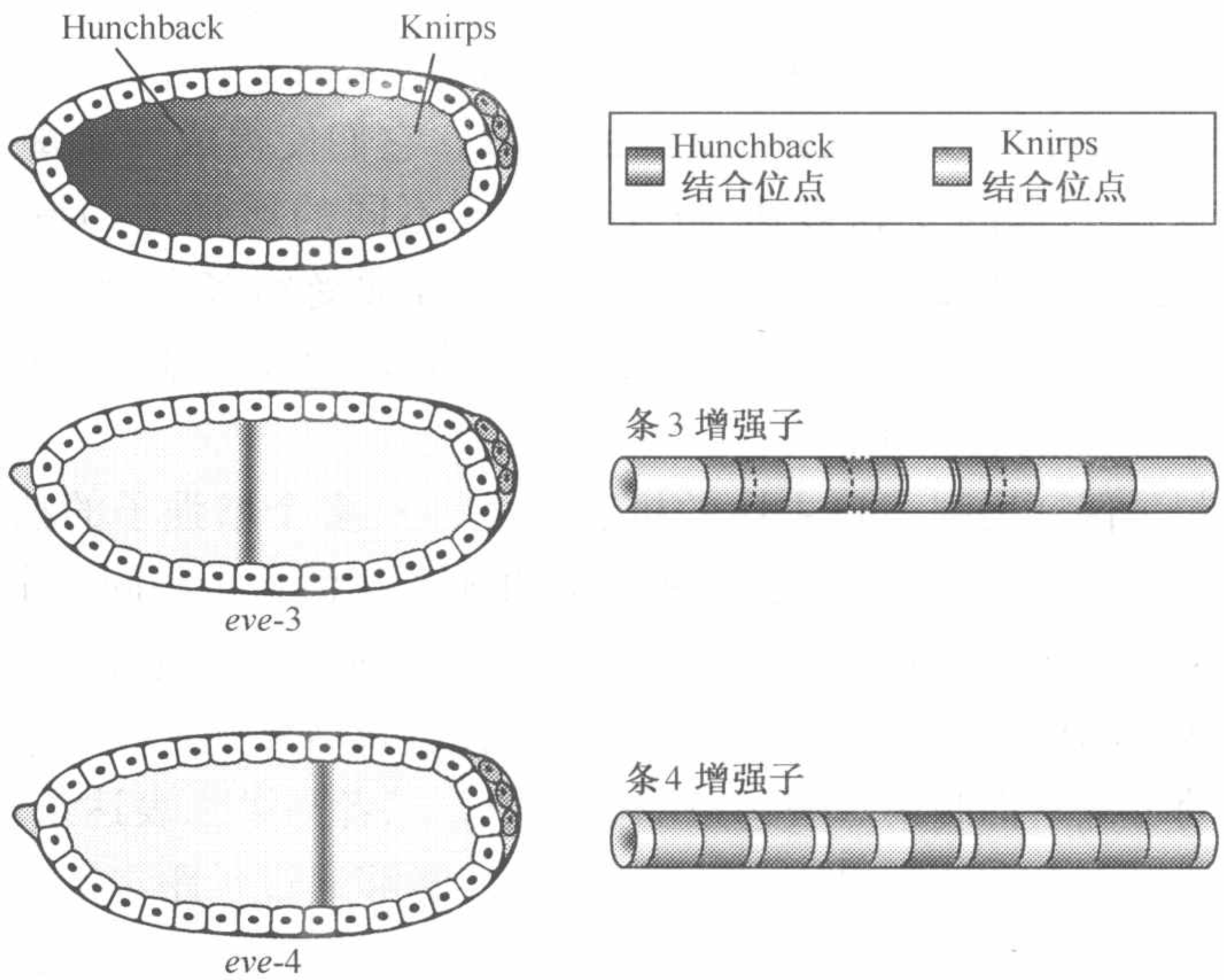


图 19-22 Hunchback 和 Knirps 反向浓度梯度对第 3 和第 4 表达条增强子的差异调节。这两个表达条存在于胚胎的不同区域：*eve* 第 3 表达条增强子分别受到高浓度的 Hunchback 和低浓度的 Knirps 抑制。相反，*eve* 第 4 表达条增强子分别受到低浓度的 Hunchback 和高浓度的 Knirps 抑制。*eve* 第 3 表达条增强子只有少数的 Hunchback 结合位点，其结果，需要高浓度的 Hunchback 使其受到抑制。*eve* 第 3 表达条增强子包含多个 Knirps 结合位点，因而，低浓度的 Knirps 就可使其受到抑制。*eve* 第 4 表达条增强子具有相反的抑制子结构，它包含大量的 Hunchback 结合位点，因此，低浓度的 Hunchback 就可抑制第 4 表达条的基因表达。第 4 表达条增强子仅有少数的 Knirps 结合位点，只有高浓度的 Knirps 才能使其受到抑制。需要注意的是，第 3 表达条增强子指导两个表达条的基因表达，第 3 条和第 7 条；第 4 表达条增强子指导第 4 条和第 6 条的基因表达。

抑制 *eve* 第 4 表达条增强子（图 19-22）。Hunchback 抑制子浓度梯度对两种增强子的差异调节产生了第 3 和第 4 表达条基因表达类型的特有的前界。Knirps 蛋白在前细胞胚胎的分布也存在浓度梯度，与第 3 表达条增强子相比需要更高的浓度来抑制第 4 表达条增强子，这种差别造成了第 3 和第 4 表达条基因表达类型的不同的后界。

我们已经见到了 Hunchback 抑制子浓度梯度产生的 Krüppel、Knirps、Giant 不同类型的基因表达，这种差异调节可能与 Krüppel、Knirps、Giant 增强子增加的 Hunchback 结合位点有关，类似的原理也适用于 Hunchback 和 Knirps 浓度梯度对第 3 和第 4 表达条增强子的差异调节。*eve* 第 3 表达条增强子包含相对较少的 Hunchback 结合位点和相对较多的 Knirps 结合位点。*eve* 第 4 表达条增强子包含相对较多的 Hunchback 结合位点，相对较少的 Knirps 结合位点（图 19-22）。类似的原理又可能操纵着剩余的表达条 *eve* 基因表达类型的调节。

短程转录抑制子允许在复杂的 *eve* 调节区域中不同的增强子独立作用

我们已经见到 *eve* 基因表达在早期胚胎受到 5 种不同的增强子的调节。事实上，在

更成熟的胚胎的心脏和 CNS 中还有更多的控制 *eve* 基因表达的增强子。这类复杂的调节并不仅仅见于 *eve* 基因，有些遗传位点会包含更多的增强子，分布于更广泛的区域。例如，在下一章我们会讨论 *homoeotic* 基因的调节，它们使得成体果蝇的体节在形态上彼此不同。数个这样的基因受到多达 10 个或更多不同的增强子的调节，它们散布在 100kb 的区间。因此，与重要发育过程相关的基因常常受到多个增强子的调节。那么，这些增强子是如何相互独立作用来产生叠加性的基因表达呢？如 *eve* 基因，5 个不同的增强子生成了 7 个不同的表达条。

短程转录抑制是保证增强子自主性的机制之一——多个增强子的独立作用产生基因表达的叠加效应，即抑制子与一个增强子的结合并不干扰活化子与同一个基因调节区域的另一个增强子的结合。例如，我们已经见到过 Krüppel 抑制子与第 2 表达条增强子结合，它建立了第 2 表达条的后界。Krüppel 抑制子第 2 表达条增强子在 500bp 附近起作用，它并不抑制核心启动子也不抑制第 3 表达条的活化子，二者与第二表达条增强子的 Krüppel 结合位点的距离超过 1kb（图 19-23）。假设 Krüppel 能够通过长距离起作用，那它就会干扰 *eve* 第 3 表达条的基因表达，因为在胚胎同一区域的 Krüppel 具有高浓度。

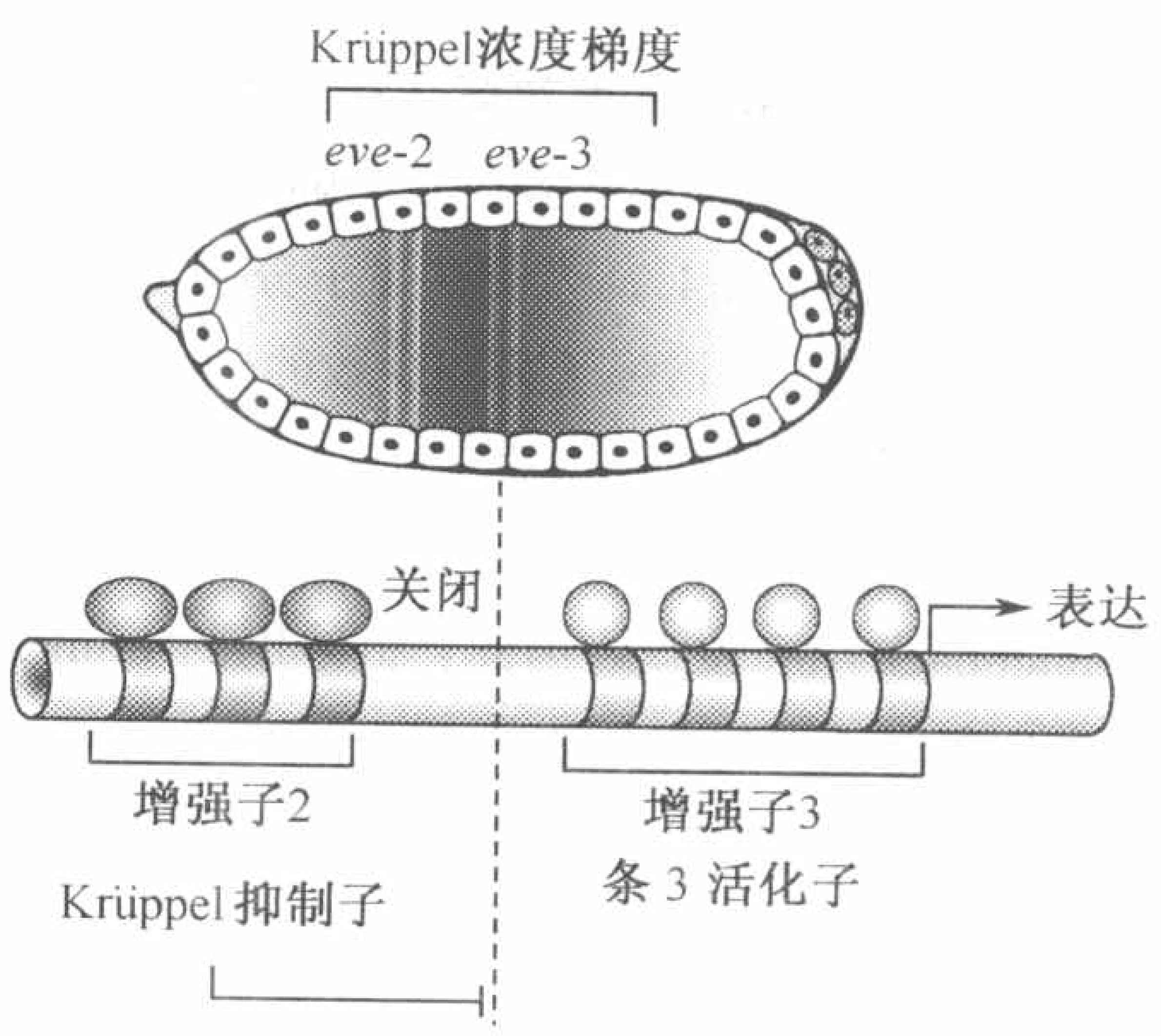


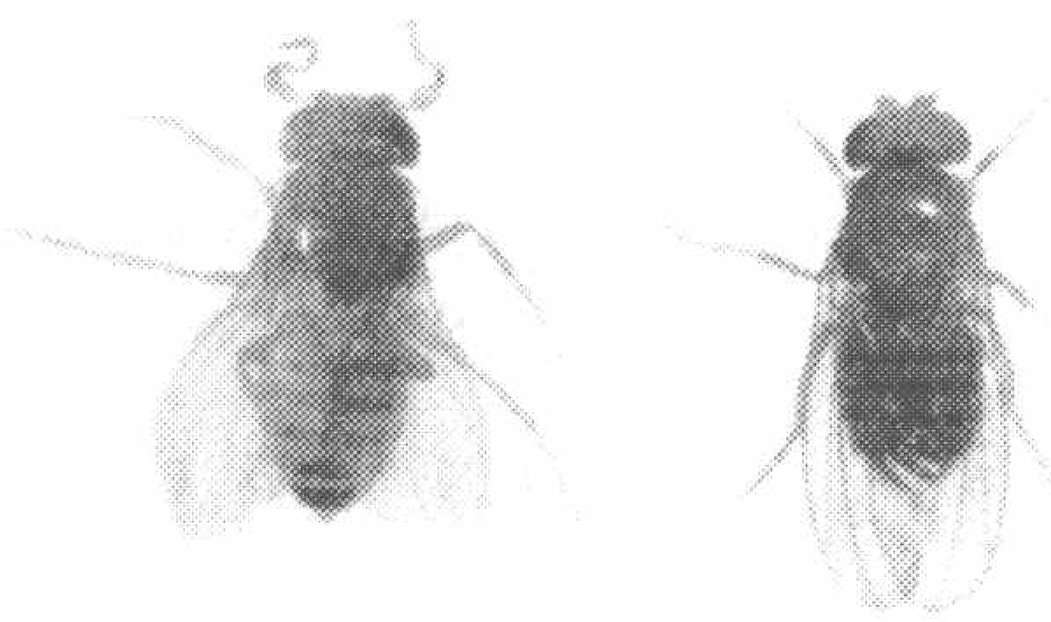
图 19-23 短程抑制和增强子自主。由于短程转录抑制，*eve* 基因调节区域不同的增强子独立发挥作用，结合于一个增强子的抑制子并不干扰邻近增强子的活化子。例如，Krüppel 抑制子结合于第 2 表达条增强子使胚胎中间区域第 2 表达条关闭。*eve* 第 3 表达条增强子在这些区域表达，它并不受 Krüppel 的抑制因为它没有 Krüppel 的特定的识别序列。此外，Krüppel 第 2 表达条增强子的结合并不干扰第 3 表达条增强子的活化子，因为它们相距甚远。Krüppel 必须在上游活化子的 100bp 距离之内结合来阻断它激活转录的能力，第 2 表达条和第 3 表达条被 1.5kb 的间隔序列分开。

同源异型基因：一类重要的发育调节因子

果蝇发育的遗传分析导致了一类重要的调节基因即同源异型基因的发现，它们造成不同体节的形态多样化。某些同源异型基因控制头节的口器和触角的发育，另一些则控制胸节的翅膀和平衡棒（halter）的形成。Antp 和 Ubx 两个研究最为透彻的同源异型基因，它们分别抑制触角和翅膀的发育。

Antp 控制胸节中部体节的发育。中胸节生成一对足其形态与前足和后足不同。Antp 为同源异型蛋白编码，通常在发育中的胚胎的中胸节表达。该基因不会在发育中的头部组织表达。然而，一个由染色体反转造成的 Antp 显性突变使其受到外来调节基因的控制，转而在头部包括触角组织表达（图 19-24）。当 Antp 在头部组织错误表达时，个体形态发生显著改变，长触角的位置长出了足。

图 19-24 Antp 的显性突变引起触角同源异型转化为足。右边的果蝇为正常的。注意头前部的初始的触角。左边的果蝇是 Antp 的显性突变杂合子（Ant-pD/+）。它可以完全成活，形态基本正常，只是在触角处伸出一对足来。



Ubx 编码一种控制第三胸节后胸节发育的同源异型蛋白。Ubx 特异性的抑制那些第二胸节即中胸节发育所需基因的表达。Antp 是其调节的基因：Ubx 抑制 Antp 在中胸节的表达，限制它在发育中的胚胎后胸节的表达。缺失 Ubx 抑制子的突变体表现出 Antp 的异常表达。此时 Antp 不仅在其正常发育中的中胸节表达还会在发育中的后胸节异常表达。这种异常表达造成后胸节转化成重复的中胸节。

在成年果蝇，中胸节拥有一对足和一对翅（图 19-25），后胸节拥有一对足和一对平衡棒。平衡棒明显小于翅，其作用是在飞行时的平衡结构。Ubx 突变体表现出惊人的表现型；拥有四只翅其中两只转化来自于平衡棒。这种突变体表现型的部分原因是 Antp 的异常表达。

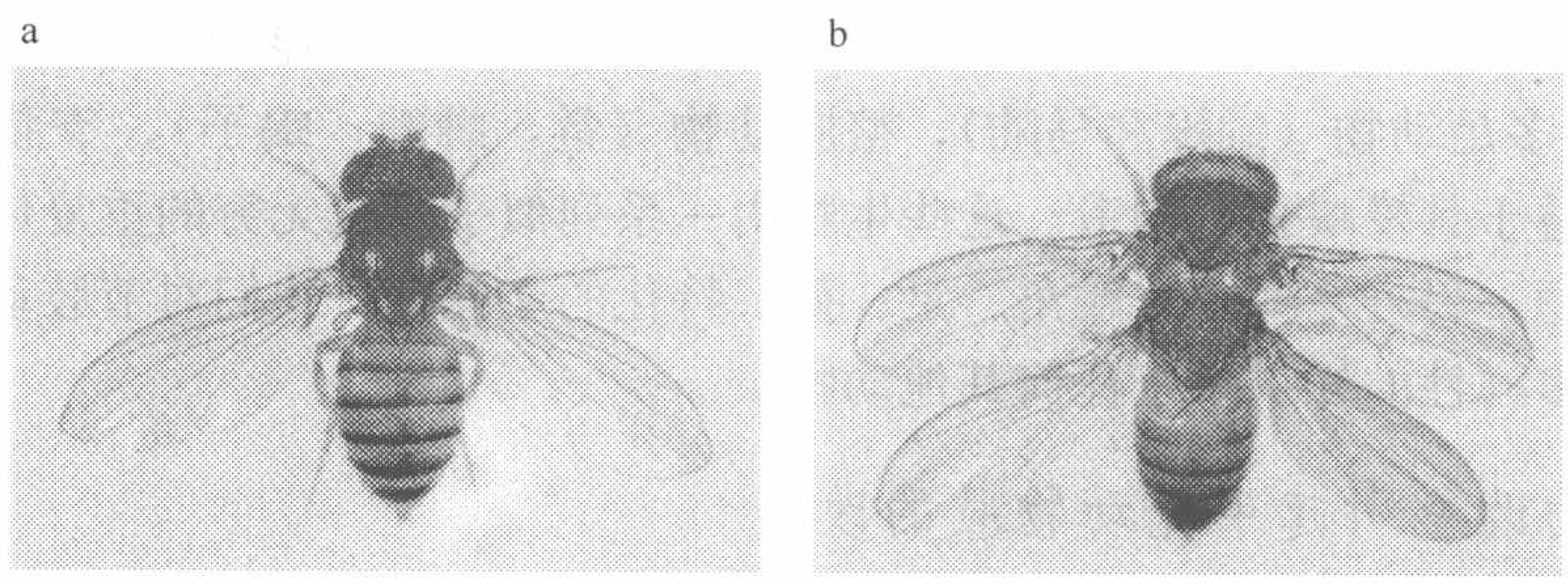


图 19-25 Ubx 的突变将后胸节转化为两个中胸节。（a）正常果蝇拥有一对主翅和后边小一些的一对平衡棒。（b）图示突变体为 Ubx 基因弱突变纯合子。后胸节转化为了重复的中胸节而不是原先的一对翅和一对平衡棒。

Ubx 在后胸节不同组织的表达依赖于一段 80 kb 的基因组调节序列。这个突变叫作 Cbx，它打断了这段调节序列但不改变该基因的蛋白质编码序列。Cbx 突变维持 Ubx 在其正常的后胸节的表达外还导致 Ubx 在中胸节的异常表达（图 19-26）。Ubx 抑制 Antp 以及其他在中胸节发育必要的基因的表达。其结果，中胸节转化成为复制的正常的后胸节。其表现型也非常不同：翅变成了平衡棒，而 Cbx 突变体看着更像是无翅的蚂蚁。

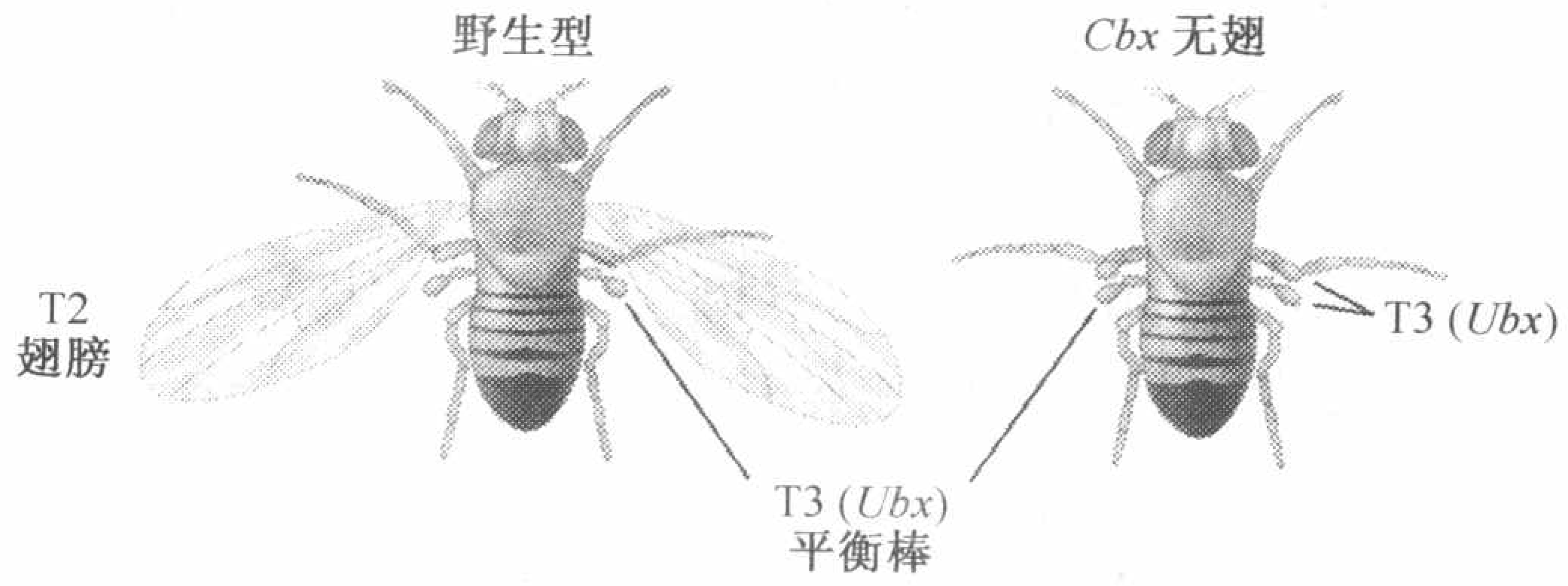


图 19-26　Ubx 在中胸节的异常表达导致翅的丢失。 Cbx 突变打断了 *Ubx* 基因的调节区域，造成它在中胸节的异常表达，导致其转化为后胸节。

同源异型基因表达的变化导致节肢动物的多样性

新生的进化-发育研究处于原本互不相干的研究领域进化生物学和发育生物学的前沿。推动进化-发育研究的动力来自于果蝇、线虫和其他模式生物的发育的遗传分析，它鉴定的主效基因是进化多样性的本源。同源异型基因是这些基因的极好的代表。

果蝇基因组拥有总共八个同源异型基因，形成两个基因簇或复合体：Antennapedia 复合体和 Bithorax 复合体（框 19-8 果蝇同源异型基因由特殊染色体簇构成）。典型的无脊椎动物基因组包含 8～10 个同源异型基因，通常分布在一个复合体内。脊椎动物复制了祖先 Hox 复合体，包含四个簇。单个同源异型基因表达和功能的改变是节肢动物肢和脊椎动物中轴骨骼变化的原因。

节肢动物变化多端

节肢动物包含五类群：三叶虫（已经灭绝），六足类（如昆虫），甲壳类（虾，龙虾，蟹等），多足动物（蚰蜒和马陆），螯肢动物（鲎，蜘蛛，蝎子）。节肢动物的成功演化部分得益于其模块化的构造。这些生物由一系列似乎变化无穷的重复体节组成。一些体节拥有翅，另外一些则有触角、足、螯，特化的交配器。我们对节肢动物的进化过程所产生的多样性的知识多于任何其他动物。

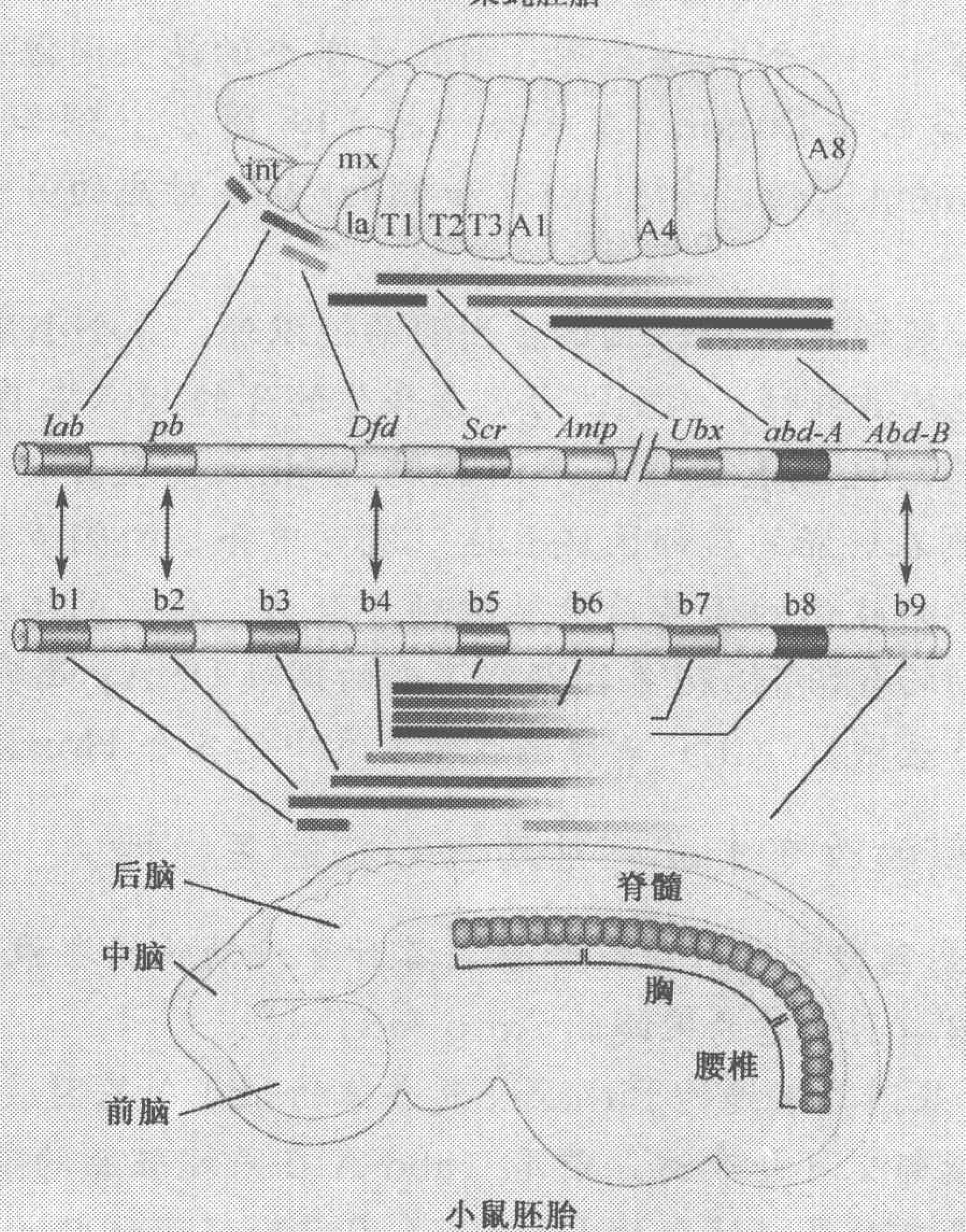
Ubx 表达的改变引起甲壳动物肢的改变

甲壳动物包括大部分游泳的节肢动物。它们部分生活在海洋中，部分生活在淡水中。它们提供了我们餐桌上的美食，如虾、蟹、龙虾。一类相当普遍用于研究的甲壳动物叫做丰年虾，又叫海猴。它的胚胎停止发育形成坚硬的孢子可从宠物店买到。其孢子加入海水后恢复发育。

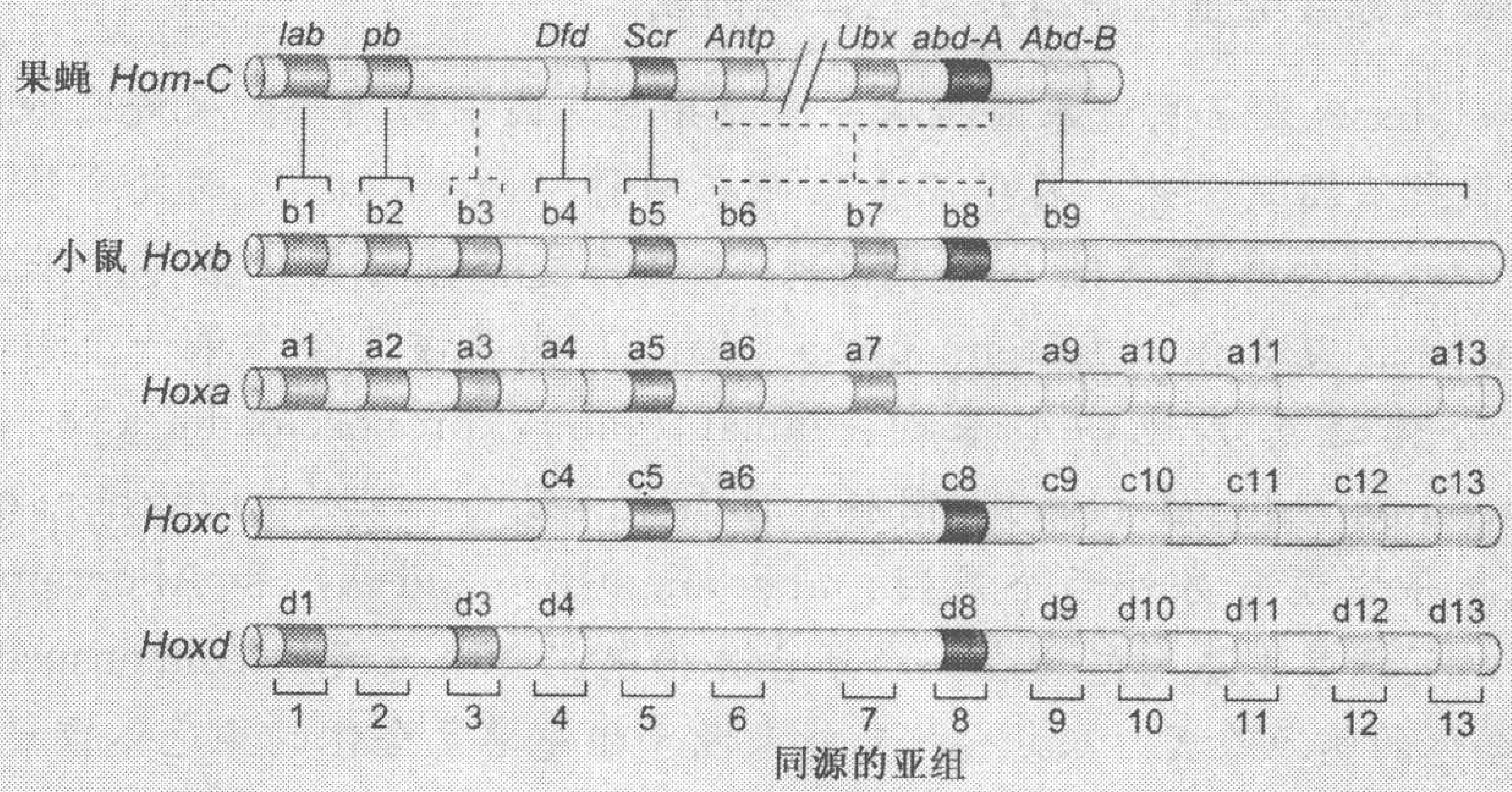
这些虾的头部拥有摄食的附肢。靠近头部的胸节，T1，包含游泳附肢，它们看起来与后面的其他胸节相像（胸节 T2-T11）。丰年虾属于甲壳动物的一个目叫作鳃足目。另一目甲壳动物叫作等足目，其第二至第八胸节为与鳃足目相似的游泳肢。但其第一胸节发生了改变。它变小了，其功能是摄食。

框 19-8 果蝇同源异型基因由特殊染色体簇构成

Antp 和 Ubx 代表了果蝇基因组八个同源异型基因中的两个。果蝇同源异型基因分布在两个簇里或基因复合体。其中五个分布在叫作 Antennapedia 的复合体中，另三个分布在 Bithorax 复合体中（框 19-8 图 1）。不要将复合体与其中的单个基因混淆。比如，Antennapedia 复合体以 Antennapedia（Antp）基因该复合体第一个被鉴定的基因命名。该复合体还有其他四个基因：labial（lab），proboscipedia（pb），Deformed（Dfd），Sexcombs reduced（Scr）。同样，Bithorax 复合体以 Ultrabithorax（Ubx）基因命名，此复合体还有另外两个基因：abdominal-A（abd-A）和 Abdominal-B（Abd-B）。面象虫的同源异型基因包含在一个复合体内包括八个果蝇的 Antennapedia 复合体和 Bithorax 复合体的同源基因。双复合体可能来源于祖先单复合体的染色体重排。同源异型基因在发育中的胚胎的前后轴向上存在着染色体排列上和基因表达上的共线性关系（框 19-8 图 1）。例如 lab 基因分布在 Antennapedia 的复合体的最 3′端就表达在果蝇发育期胚胎头部最前部。相反，Abd-B 基因分布在 Bithorax 复合体的 5′端就在最后端相应区域表达（框 19-8 图 1）。这种共线性的意义还不明确，但是应该十分重要，因为它在所有的节肢动物（包括面象虫）以及现已研究的脊椎动物包括小鼠和人类中都是保守的。



框 19-8 图 1　果蝇和小鼠的结构和表达。 图里比较了果蝇和小鼠 Hox 基因的共线性序列和表达。（经允许，引自 McGinnis W. and Krumlauf R. 1992. Cell 68: 283—302，Fig. 2. 版权所有 Elsevier）



框 19-8 图 2 果蝇和小鼠同源异型基因结构和表达的保守。（经允许，引自 Gillbert S. E. 2000. Developmental Biology，6th ed.，Fig. 11. 36. 版权所有 Sinauer）

哺乳动物 Hox 基因复合体控制前后轴类型

小鼠拥有 38 个 Hox 基因，分布在四个基因簇里（Hoxa，Hoxb，Hoxc，Hoxd）。每个簇或复合体拥有 9～10 个 Hox 基因，对应于单个的昆虫同源异型基因簇包括果蝇的 Antennapedia 复合体和 Bithorax 复合体（框 19-8 图 2）。例如，Hoxa-1 和 Hoxb-1 与果蝇的 lab 基因关联性最强，而 Hoxa-9 和 Hoxb-9 与对应的另一端的果蝇的 Abd-B 非常相似。

除去这种小鼠和果蝇的 Hox 基因的系列同源性以外，每个小鼠 Hox 基因还表现出了果蝇中的同样类型的共线性。例如，每个复合体的 3′的 Hox 基因如 Hoxa-1 和 Hoxb-1 在发育胚胎的最前端区域表达（未来的后脑）。相反，复合体 5′端的 Hox 基因如 Hoxa-9 和 Hoxb-9 则在胚胎的后部区域表达（发育中脊椎的胸节和腰节）。Hoxd 复合体表现出了在发育中的肢体的沿着前后轴的顺序表达。相应的类型并没有在昆虫中出现，提示随着脊椎动物金华 Hoxd 基因获得了新的调节 DNA。事实上，在第 17 章已经讲到特化的全面调节区域（GCR）来协调肢体发育中的各个 Hoxd 基因的表达。

Hox 基因表达类型的改变生成脊椎动物形态的多样性

哺乳动物 Hox 基因突变造成中轴骨骼（各种脊椎动物的脊椎或脊柱）的中断。这些形态变化令人想起前面讲过的果蝇的 Antp 和 Ubx。

小鼠的 Hoxc-8 基因与果蝇的 Bithorax 复合体的 abd-A 基因最为相像。它正常情况下在脊柱的肋骨和腰部交界处表达。前尾（abd-A 在果蝇胚胎的前腹部表达）。最初的腰型脊椎动物通常没有肋骨。纯合的 Hoxc-8 基因敲除突变体胚胎表现出显著的突变表现型。第一个腰椎发育出额外的发育不全的肋骨。这种发育的异常现象还被称为同源异型转化，一种正常结构长在了错误的位置。此种情况下，通常后位的胸骨发育在了前位的腰椎区域。

Hox 基因表达模式的保持

在果蝇和小鼠胚胎早期，Hox 基因局部的表达模式是由序列特异的转录激活和抑制共同作用来建立的。一些调控蛋白被整合到模块化的细胞信号途径中，如 FGF 和 Wnt 信号途径。在果蝇中，许多负责 eve 基因的局部表达的 Gap 抑制子也控制着起始阶段 Hox 基因的表达。这些表达模式在整个生命周期中都将被保持，即使是 Gap 抑制子丧失后很久。

让我们来看一个例子，果蝇中的 Abd-B 基因。该基因特异的在后腹部表达，包括第 5-8 腹节。在果蝇胚胎发育的早期的头、胸和前腹中，Abd-B 的表达最初被 Hb、Kr 和 Kni Gap 抑制子抑制。这是维持 eve 局部条纹化表达的同样一批抑制子（图 19-29，图 19-20）。这些抑制子限制了 Abd-B 在腹节前部的表达。

Abd-B 基因表达的维持以及大部分其他 Hox 基因的表达依赖于一个大的叫作多梳抑制复合体和蛋白质复合体（PRC）。PRC 与那些在早期胚胎没能活化的 Abd-B 基因的调节序列结合：包括头部，胸部，前腹部的前体。在所有这些细胞，PRC 造成组蛋白 3 的 27 位赖氨酸甲基化，此甲基化与 Abd-B 转录单位的抑制有关。相对应的，一个普遍存在的活化复合体三胸复合体（TRC）与早期胚胎细胞（腹节后部）的活化的 Abd-B 基因的调节序列结合。TRC 的结合造成组蛋白 3 的 4 位赖氨酸甲基化，这与 Abd-B 基因的活化相关联。

因此，取决于最初的基因表达情况，PRC 和 TRC 在早期胚胎维持 Hox 基因的开和关的状态。如果特定的 Hox 基因在特定的细胞被抑制，PRC 结合并保持其所有的子代细胞该基因的关闭。相对应的，如果特定的 Hox 基因在特定的细胞被活化，TRC 结合并保持其所有的子代细胞该基因的稳定表达，PRC 起到了调节记忆 Hox 基因表达类型的作用。

微小 RNA 调节 Hox 基因活性

许多 Hox 基因复合体包含微小 RNA 基因（microRNA 或 miR）。比如，果蝇 ANT-C 含有 miR-10，BX-C 含有 miR-lab-4。这些编码的微小 RNA 抑制或减弱不同的 Hox 蛋白质的合成，lab-4 miRNA 抑制腹部 Ubx 蛋白质合成。脊椎动物 Hox 基因复合体也含有 miR 基因，包括 miR-10。miR-196 基因分布在多个脊椎动物 Hox 基因复合体的 5′端，其编码的 miRNA 据认为抑制小鼠胚胎后部的 Hoxb8 蛋白质的合成。参见第 18 章关于 miRNA 如何阻止或减弱蛋白质合成。

腮足动物和等足动物的 Ubx 基因表达类型稍有不同。这些不同与等足动物第一胸节的游泳肢的修饰有关。也许，腮足动物和等足动物的共同祖先拥有像丰年虾那样的胸肢（其本身也是腮足动物的一员）：其胸节都包含游泳肢。在腮足动物和等足动物的分化中 Ubx 的调节序列在等足动物发生改变。其结果在第一胸节的 Ubx 基因表达被消除，只局限在 T2-T8 胸节。不难想象 Ubx 抑制了胸节一或数个头类型基因的表达。在

丰年虾，这些头基因在所有的 11 个胸节都是关闭的，然而在等足动物，由于失去了 Ubx 的抑制子，头基因在 T1 胸节表达。确实，Scr 基因表达局限在腮足动物的头部但在等足动物是在 T1。Scr 在 T1 的表达造成螯肢的发育而不是通常的游泳肢（图 19-27）。

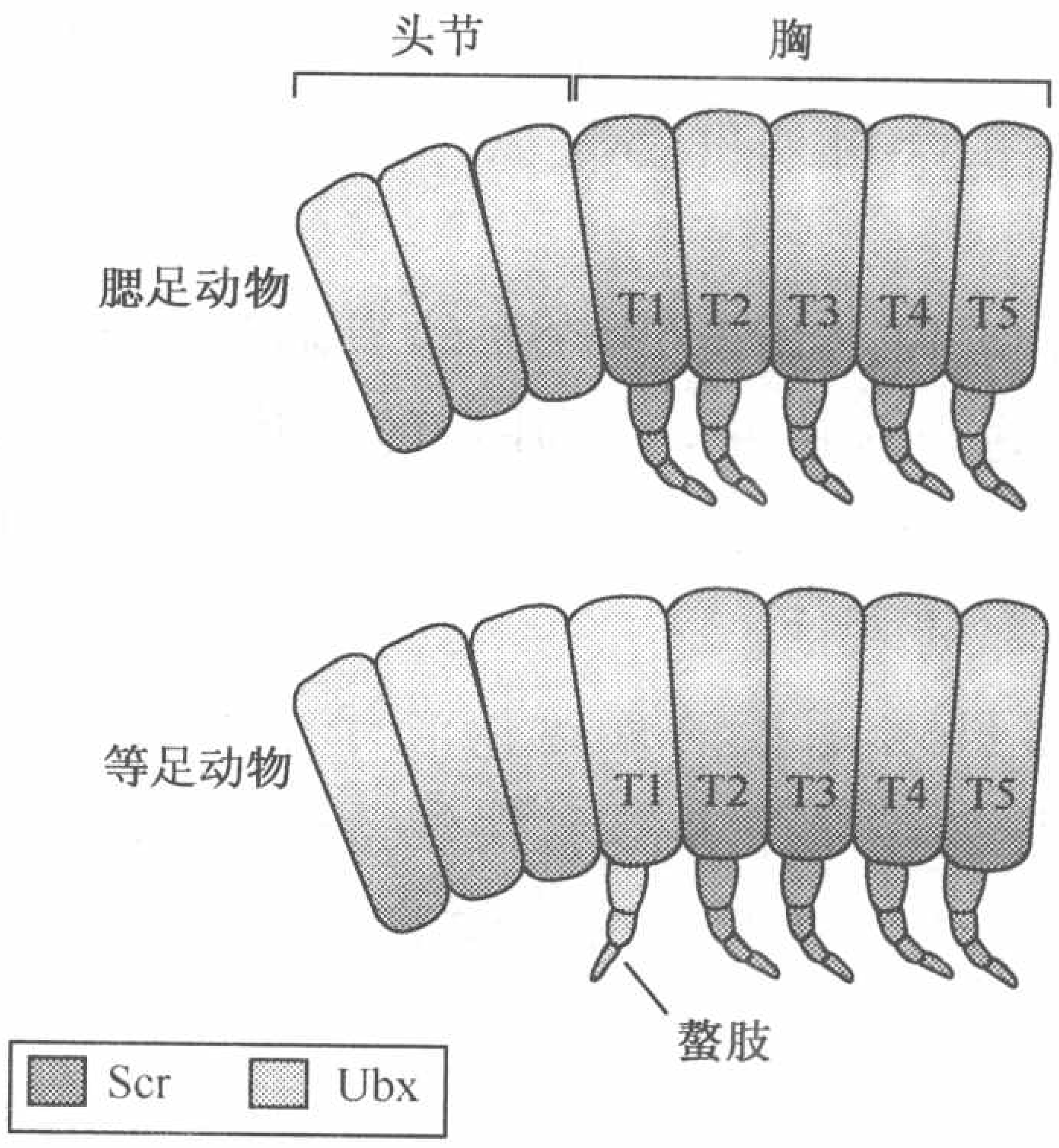


图 19-27 两组不同甲壳动物的形态变化。 在腮足动物，Scr 基因表达局限在头部有助于摄食附肢的发育，而 Ubx 在胸节的表达则控制游泳肢的发育。在等足动物，Scr 基因的表达同时可在头部和第一胸节（T1）检测到，其结果，T1 的游泳肢转化成了摄食附肢（螯肢）。Scr 的向后扩展是 T1 胸节失去 Ubx 的表达的结果，而 Ubx 通常会抑制 Scr 的表达。

Ubx 基因表达在腮足动物和等足动物的不同的基础是什么呢？有多种不同的可能的解释，最为可信的是等足动物的 Ubx 基因的调节序列发生了突变。在这个模型里，Ubx 增强子不再介导其在 T1 胸节的表达。事实上，甲壳动物胸节的 Ubx 表达的缺失与摄食附肢的发育存在紧密的关联性。例如龙虾胚胎的头二个胸节缺乏 Ubx 基因表达具有两对螯肢。清道夫虾的头三个胸节缺乏 Ubx 基因表达具有三对螯肢。

为什么昆虫缺少腹肢

所有昆虫有六条腿，三个胸节上每节上有一对，这在一百多万种昆虫中无一例外。相反，其他节肢动物如甲壳动物则拥有数目不等的肢。有些甲壳动物在每节胸节和腹节上都有肢。这种进化上的形态变化，比如昆虫腹节失去肢并非是因为类型决定基因的表达改变，如在等足动物出现的那样。事实上，昆虫腹肢的失去是因为 Ubx 调节蛋白质的功能的改变。

在昆虫，Ubx 和 abd-A 抑制一个关键的肢节发育必需的叫作 Distal-less（Dll）的基因的表达。在发育的果蝇胚胎，Ubx 在后胸和前腹大量表达；abd-A 的表达延展到更靠后的腹节部位。总而言之，Ubx 和 abd-A 使得 Dll 在头七个腹节保持关闭状态。尽管 Ubx 也在后胸节表达，它并不干扰 Dll 在那里的表达，因为只是在 Dll 基因活化后才在这个 T3 腿部位表达。其结果，Ubx 并不干扰 T3 肢的发育。

在甲壳动物如上述的腹足动物丰年虾在所有 11 个胸节都有大量的 Ubx 和 abd-A 的表达。Dll 的表达促进游泳肢的发育。为什么 Ubx 抑制 Dll 在昆虫腹节的表达而不是在甲壳动物呢？答案是 Ubx 蛋白质在昆虫和甲壳动物之间已经发生了分化。下述实验能

够证明这一点。

Ubx 基因在转基因果蝇胚胎预期的胸节的失误表达抑制了 Dll 的表达进而抑制了肢的发育（图 19-28）。相反，甲壳动物 Ubx 基因在转基因果蝇胚胎预期的胸节的失误表达并不像 Dll 的表达和肢的形成。这些观察表明果蝇的 Ubx 蛋白和甲壳动物的 Ubx 蛋白在功能上的差异。果蝇 Ubx 蛋白抑制 Dll 基因表达而甲壳动物 Ubx 蛋白则不能够。

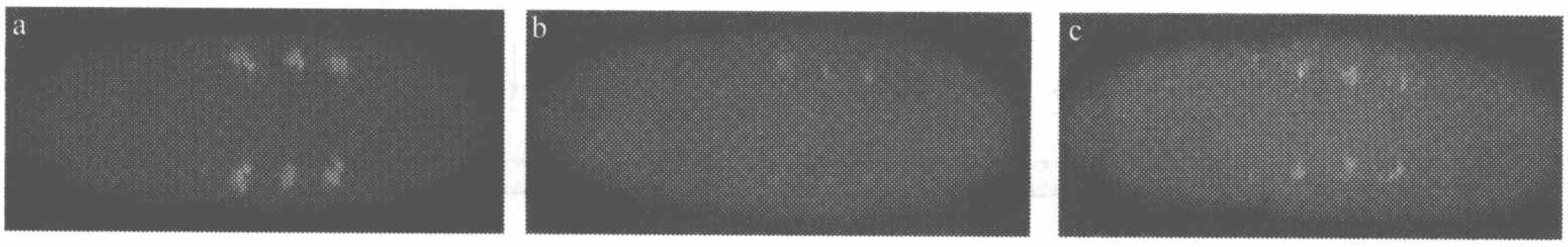
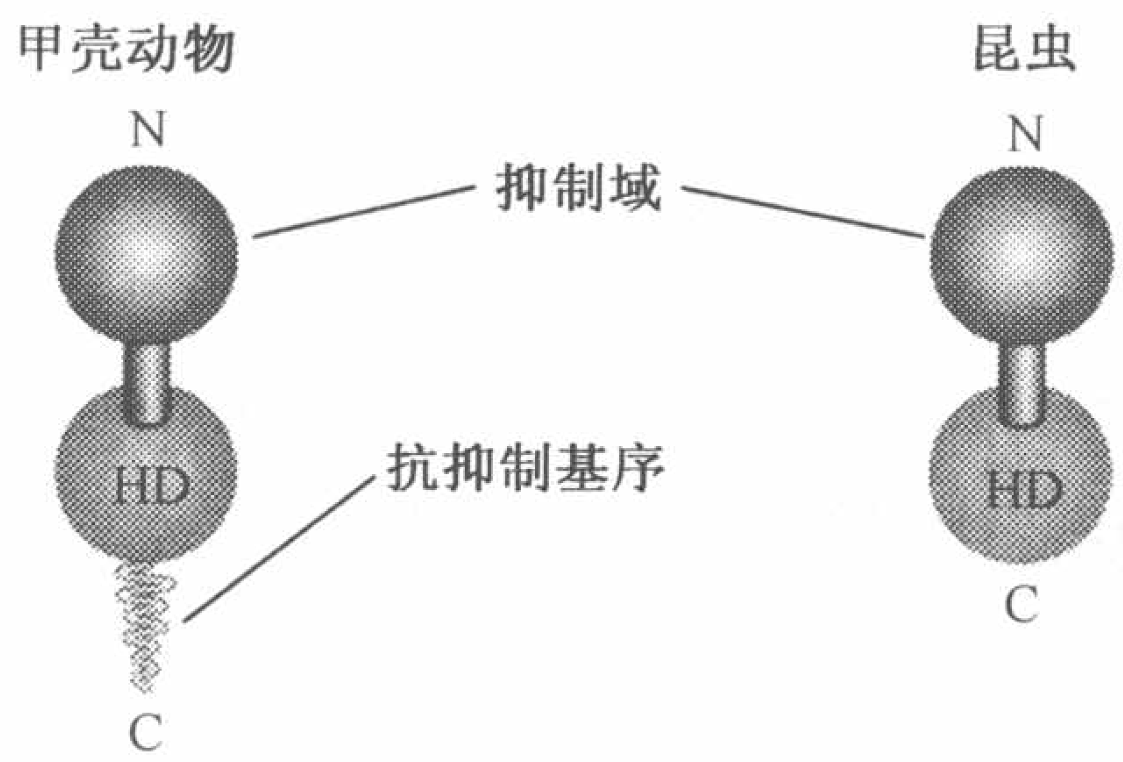


图 19-28 Ubx 蛋白功能的进化改变。(a) Dll 增强子（Dll304）通常在果蝇胚胎的三对部位被活化。这些部位形成成年果蝇的三对腿。(b) 果蝇 Ubx 蛋白（DmUbxHA）的失误表达强烈抑制了来自 Dll 增强子的表达。(c) 相反，丰年虾 Ubx 蛋白的失误表达（AfUbxHA）只造成对 Dll 增强子轻微的抑制。

那么 Ubx 蛋白的这些功能差异的基础是什么呢（它们共享 32％的氨基酸总体水平上的相似性但是它们的同源异型域却几乎完全相同——59/60 匹配）？甲壳动物的这个蛋白拥有一个具有抑制活性的 29 个氨基酸的短模体（motif）。当这个模体被剔除后，该蛋白抑制 Dll 表达的能力与果蝇的相同。

甲壳动物和果蝇的 Ubx 蛋白含有多个抑制区域。在第 17 章已经讨论过，很可能这些区域与一个或多个转录抑制复合体发生反应。甲壳动物 Ubx 蛋白的抗抑制肽可能干扰抑制区域对募集这些抑制复合体的能力。当这个肽接到果蝇蛋白上，这个杂合蛋白像甲壳动物一样不再抑制 Dll。

图 19-29 甲壳动物和昆虫 Ubx 蛋白的比较。(左) 甲壳动物的 Ubx 蛋白。羧基末端的抗抑制肽封闭了氨基端抑制区域的活性。(右) 昆虫的 Ubx。突变失去了羧基末端的抗抑制肽。



飞行肢的修饰可能来自调节 DNA 序列的进化

Ubx 主导了我们对于节肢动物形态变化的讨论。Ubx 基因表达的变化似乎与甲壳动物的游泳肢转化成螯肢有关。更进一步，Ubx 蛋白失去抗抑制肽也能解释昆虫腹肢的退化。在这一节，我们将综述 Ubx 目标基因调节序列改变的能够解释果蝇和蝴蝶翅形态变化的证据。

在果蝇中，Ubx 在发育中的平衡棒中表达，其作用是抑制翅的形成。大约 5～10 个基因被 Ubx 所抑制。这些基因与翅的生长和类型至关重要的基因（图 19-30）都在发

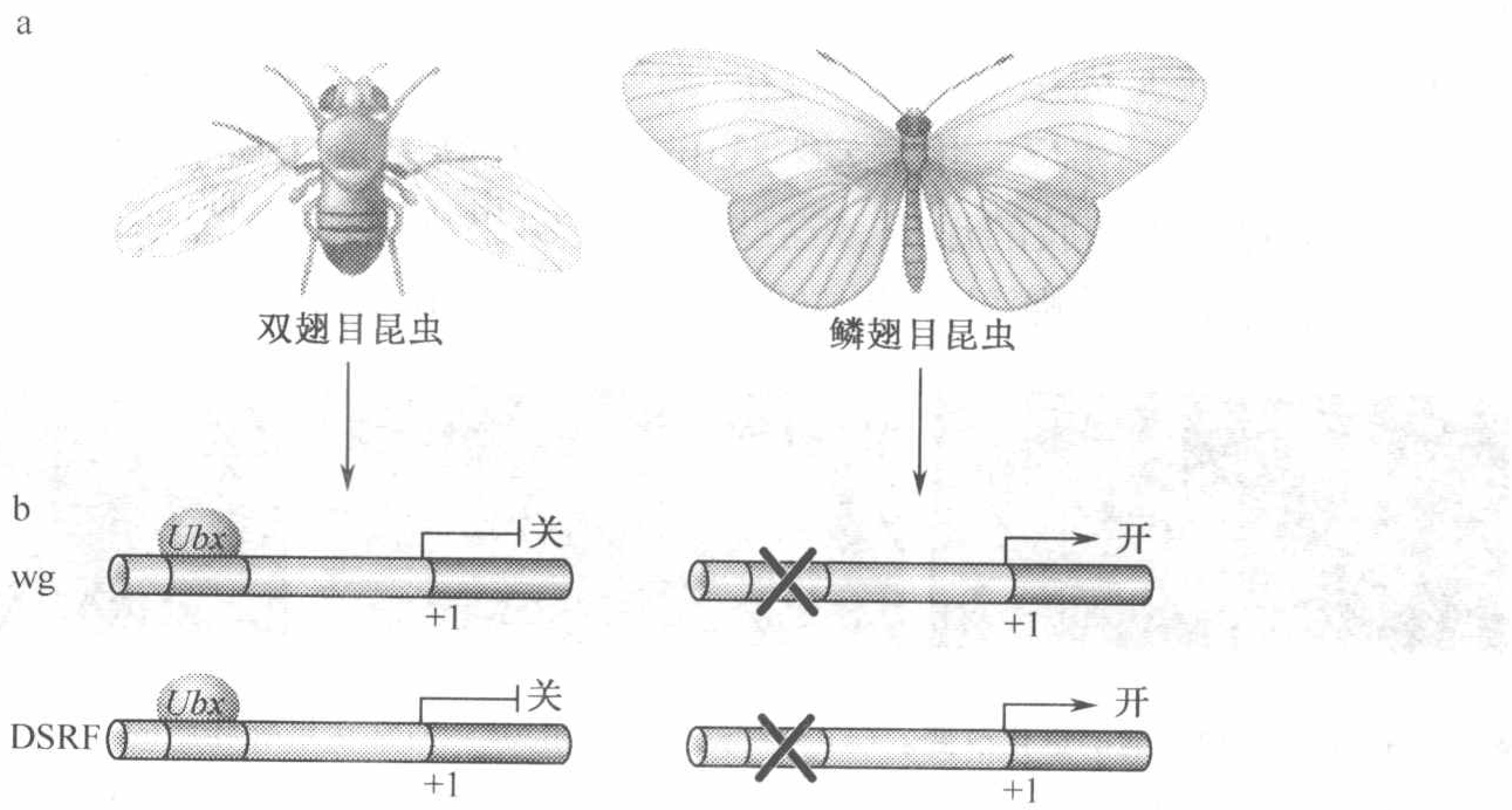


图 19-30　Ubx 靶基因调节 DNA 的变化。（a）Ubx 抑制子在双翅目动物的平衡棒和鳞翅目动物的后翅中表达（橙色）。（b）在双翅目动物，不同的靶基因拥有 Ubx 抑制子位点，而在鳞翅目则失去了这些位点。

育的翅中表达。在 Ubx 突变体中这些基因在平衡棒中不再被抑制，其结果，平衡棒发育成了第二对翅。

果蝇是双翅目动物，它们都拥有一双翅和一对平衡棒。有可能 Ubx 在所有的双翅目中起到了翅发育的抑制作用。蝴蝶属于昆虫纲的不同的目，鳞翅目。所有这个目（包括蛾子）拥有两对翅而不是一对翅和一对平衡棒。这些双翅目和鳞翅目昆虫翅形态的差异的基础是什么呢？

这两个目的共同祖先约在 250 万年前分化。大约是在人类和非哺乳类如青蛙分化的时间。这个期间给出了足够的时间用以上讨论的三种策略去改变 Ubx 基因的功能。最简单的机制是改变 Ubx 基因表达类型因而它在鳞翅目昆虫的后翅前体细胞失去了表达。这样发育中的后翅就可以表达所有被 Ubx 抑制的基因。为等足动物游泳肢向螯肢的转化提供了清晰的例证。然而，果蝇和蝴蝶的 Ubx 基因表达类型并无明显改变；Ubx 在整个蝴蝶发育中的后翅大量表达。

这为我们带来两种可能性。第一，在果蝇和蝴蝶里 Ubx 蛋白功能不同。第二就是 5～10 种受 Ubx 抑制的每个基因的调节 DNA 发生了改变，它们不再受到 Ubx 的抑制（图 19-30）。看来，更易于改变 Ubx 的抑制活性较之于改变 Ubx 靶基因的调节序列。

令人惊异的是不大可能的解释——多个 Ubx 靶基因调节序列的改变——成为了不同翅形态变化的合理解释。Ubx 蛋白看来在果蝇和蝴蝶行使了相同的功能。例如，在蝴蝶 Ubx 后翅部分细胞的缺失导致它们转化为前翅结构（图 19-30a 前翅和后翅的差别）。这个观察表明 Ubx 蛋白在前翅发育中起到了抑制发育的抑制子的作用。虽然还没有证明，但是非常可能翅类型基因的调节 DNA 失去了 Ubx 结合位点（图 19-30b）。因而，它们在发育的后翅不再受到 Ubx 的抑制。

小结

发育中的胚胎细胞表达不同组的基因，因而遵循不同的发育途径。大部分差异基因表达在转录起始水平上受到调节，这包括 3 个策略：mRNA 定位、细胞-细胞接触、分泌信号分子的扩散。

mRNA 的定位是通过其特定的 3′-UTR 序列与微管的生长端的结合来实现的。这个机制被用于 *ash1* mRNA 定位于出芽酵母的子细胞，它还被用于 *oskar* mRNA 定位于果蝇未受精卵的后端细胞质。

在细胞-细胞接触中，与细胞膜结合的信号分子通过活化相邻细胞信号途径改变基因表达，在某些情况下，静止的转录活化子或者共活化子蛋白从细胞表面释放进入细胞核。还有一些情况，静止的转录因子（或者转录抑制子）已经存在于核内，它们受到修饰后参与基因表达的活化。细胞-细胞接触方式被枯草芽孢杆菌用来建立母细胞和前孢子不同的基因表达程序。非常类似的机制存在于昆虫 CNS 发育时阻止表皮细胞形成神经元。

分泌的细胞信号分子的细胞外浓度梯度在复杂组织或器官的发育中建立了多种细胞类型。这些浓度梯度产生了活化的转录因子的细胞内浓度梯度，它在进一步以浓度依赖的形式控制基因表达。细胞外 Sonic Hedgehog 浓度梯度在脊椎动物神经管的腹部一半形成 Gli 活化子的浓度梯度。不同浓度的 Gli 调节不同种类的靶基因，生成不同的神经细胞类型。同样，早期果蝇胚胎的 Dorsal 浓度梯度引起横跨背-腹轴的不同的基因表达类型，这种基因表达的差异调节依赖于确定的增强子对 Dorsal 结合的亲和力。

果蝇胚胎的分节依赖于 mRNA 定位和调节因子的共同作用。*bicoid* 和 *oskar* mRNA 分别定位在前、后极导致形成横跨前-后轴的广泛的 Hunchback 浓度梯度。这个浓度梯度依次在未来的胸部和腹部建立了 Krüppel、Knirps、Giant 的浓度梯度。这 4 种蛋白质总称为间隔蛋白；它们起到转录抑制子的作用，建立成对出现的局部表达条。单个表达条受到成对基因，如 *eve* 的调节区域的增强子的调节。每个增强子包含活化子以及间隔抑制子的多个结合位点，广泛存在的活化子如 Bicoid 与局部的间隔抑制子相互作用建立了每个成对表达条的前后边界。不同的增强子独立发挥作用形成复合的、7 个表达条的成对的基因表达。这种增强子自主调节部分来自于短程转录抑制，一个间隔抑制子对增强子的结合并不干扰同一基因相邻表达条增强子的活性。

同源异型基因编码形成相互不同的体节的调节蛋白。两个深入研究的同源异型基因 Antp 和 Ubx 在果蝇分别控制第二和第三胸节的发育。Ubx 在发育中的翅中失误表达产生无翅果蝇，而在头部的失误表达则导致触角转化成了腿。

仅仅单从数量和多样性上讲，节肢动物是最成功的动物门类。对于节肢动物多样性的分子基础的了解远胜于任何其他动物类群。例如，Ubx 基因的表达类型的变化与各种甲壳动物的游泳肢向螯肢的转化相关。Ubx 蛋白的功能改变能够解释昆虫腹肢的抑制。最后一点，Ubx 靶增强子的变化能够解释双翅目动物的平衡棒和鳞翅目动物的后

翅的不同形态。

（刘斌 刘国振 译校）

参考文献

书

Carroll S.B., Grenier J.K., and Weatherbee S.D. 2004. *From DNA to diversity: Molecular genetics and the evolution of animal design*, 2nd ed. Blackwell, Malden, Massachusetts.

Davidson E.H. 2006. *The regulatory genome: Gene regulatory networks in development and evolution.* Academic Press, San Diego.

Gilbert S.F. 2006. *Developmental biology*, 8th ed. Sinauer Associates, Sunderland, Massachusetts.

Wolpert L., Smith J., Jessell T., Lawrence P., Robertson E., and Meyerowitz E. 2007. *Principles of development*, 3rd ed. Oxford University Press, Oxford.

mRNA 定位

King M.L., Messitt T.J., and Mowry K.L. 2005. Putting RNAs in the right place at the right time: RNA localization in the frog oocyte. *Biol. Cell* **97:** 19–33.

Rongo C. and Lehmann R. 1996. Regulated synthesis, transport and assembly of the *Drosophila* germ plasm. *Trends Genet.* **12:** 102–109.

Weil T.T., Forrest K.M., and Gavis E.R. 2006. Localization of bicoid mRNA in late oocytes is maintained by continual active transport. *Dev. Cell* **11:** 251–262.

细胞-细胞接触

Greenwald I. 1998. LIN-12/Notch signaling: Lessons from worms and flies. *Genes Dev.* **12:** 1751–1762.

Kimble J. and Simpson P. 1997. The LIN-12/Notch signaling pathway and its regulation. *Annu. Rev. Cell Dev. Biol.* **13:** 333–361.

Schweisguth F. 2004. Notch signaling activity. *Curr. Biol.* **14:** R129–R138.

Shapiro L., McAdams H.H., and Losick R. 2002. Generating and exploiting polarity in bacteria. *Science* **298:** 1942–1946.

形态梯度

Ashe H.L. and Briscoe J. 2006. The interpretation of morphogen gradients. *Development* **133:** 385–394.

Ephrussi A. and St Johnston D. 2004. Seeing is believing: The bicoid morphogen gradient matures. *Cell* **116:** 143–152.

Moussian B. and Roth S. 2005. Dorsoventral axis formation in the *Drosophila* embryo—Shaping and transducing a morphogen gradient. *Curr. Biol.* **15:** R887–R899.

分节

Clyde D.E., Corado M.S., Wu X., Pare A., Papatsenko D., and Small S. 2003. A self-organizing system of repressor gradients establishes segmental complexity in *Drosophila*. *Nature* **426:** 849–853.

Fujioka M., Emi-Sarker Y., Yusibova G.L., Goto T., and Jaynes J.B. 1999. Analysis of an *even-skipped* rescue transgene reveals both composite and discrete neuronal and early blastoderm enhancers, and multi-stripe positioning by gap gene repressor gradients. *Development* **126:** 2527–2538.

Jaeger J. and Reinitz J. 2006. On the dynamic nature of positional information. *BioEssays* **28:** 1102–1111.

同源基因

Browne W.E. and Patel N.H. 2000. Molecular genetics of crustacean feeding appendage development and diversification. *Semin. Cell Dev. Biol.* **11:** 427–435.

Krumlauf R. 1994. *Hox* genes in vertebrate development. *Cell* **78:** 191–201.

Lemons D. and McGinnis W. 2006. Genomic evolution of *Hox* gene clusters. *Science* **313:** 1918–1922.

Lewis E.B. 1978. A gene complex controlling segmentation in *Drosophila*. *Nature* **276:** 565–570.

Maeda R.K. and Karch F. 2006. The ABC of the BX-C: The bithorax complex explained. *Development* **133:** 1413–1422.

Popadić A., Abzhanov A., Rusch D., and Kaufman T.C. 1998. Understanding the genetic basis of morphological evolution: The role of homeotic genes in the diversification of the arthropod bauplan. *Int. J. Dev. Biol.* **42:** 453–461.

Weatherbee S.D., Nijhout H.F., Grunert L.W., Halder G., Galant R., Selegue J., and Carroll S. 1999. Ultrabithorax function in butterfly wings and the evolution of insect wing patterns. *Curr. Biol.* **9:** 109–115.

第20章 基因组分析与系统生物学

本章概要

• 基因组学概述

• 系统生物学

分子生物学的性质已经被技术的进展所改变。对于一些复杂的细胞学过程，譬如一个干细胞分化成为心肌的过程，现在我们可以鉴定其中每一个组分、每一个基因和每一个蛋白质。在没有大规模 DNA 测序技术和蛋白质组学的方法之前，分子生物学家只能够从全部成分中解剖其中一小部分，并用此来阐明其大致原理。从他们的角度来看，这一小部分是他们所研究的对象中关键的限速调控成分。鉴定和分析一个生物学过程中所有组分的能力，为解答这一系列新问题提供了机会：这些生物学过程的设计原理是什么？在本章中，我们将对基因组和系统生物学的一些新出现的领域进行讨论，这些新的领域是传统的基于实验的分子生物学与计算机分析结合的产物。

基因组学概述

在全基因组测序之前，要进行大尺度的 DNA 序列的比较是很困难的事情。人们最多只能对少数物种的少数基因的 DNA 序列进行研究。随着功能强大的全自动测序仪的问世，现在已经能够获得整个基因组的组织与遗传组分的完整信息。事实上，截止到本书撰写之时，超过 1000 个物种的基因组序列已经被测定并完成组装。因此，现在我们可以对各种微生物、植物和动物之间的完整遗传组分进行比较分析。在本章中，我们将首先阐述基因组注释的一些基本方法，即采用实验和计算的方法从复杂的整个基因组中鉴定每个基因（包括内含子-外显子结构）和相关的调控序列。然后，我们将讨论基于基因组的比较和各个基因组的系统分析所获得的一些新的知识。

生物信息学技术用于全基因组中蛋白质编码基因的鉴定

基因组序列由成万上亿个 A、C、T、G 按次序排列而成，它们组成了有机体的各条染色体［见第 21 章中关于基因组如何由短的“支架”（scaffolds）组装而成的详细讨论］。如果不对基因组进行“注释”，其巨大冗长的序列不能提供任何有价值的信息。如后面几页所描述的，注释（annotation）是指系统鉴定每一段基因组 DNA 中所包含的蛋白质编码序列和非编码序列，后者是指那些调控 RNA，如 miRNA 的 DNA 序列（见

第 18 章)。对于每一个转录单元，都要确定其内含子与外显子的准确结构，如果所研究的基因组与模式生物（如酵母和果蝇）接近，就有可能确定基因组中大部分基因的潜在的或已知的功能。只有获得了这些信息，才有可能把一个基因组的全部编码能力分门别类并和其他物种的基因组进行比较。

细菌和简单的真核生物的基因组注释相对直接，只要能够识别可读框（ORF）即可。尽管并非所有的 ORF 都是真正的蛋白质编码基因，尤其是那些小 ORF。可读框的识别相对而言较为容易，最具挑战的还是如何确定基因的功能。

对于具有复杂的内含子-外显子结构的动物基因组来说，这一挑战还要大得多。正因为如此，我们需要多种生物信息学分析工具来鉴定基因，从复杂的基因组中识别其遗传组分。许多计算机程序已经被开发出来。根据设定的各种序列参数，这些程序可以应用于鉴定潜在的蛋白质编码基因（图 20-1)，包括依据两侧的 5′和 3′剪切位点来确定 ORF。在第 13 章中我们已经讨论到，剪接供体和受体的序列既短又具有简并性（degenerate)，但如果我们结合前后的其他信息，譬如表达序列标签（expressed sequence tag，EST）数据的话（下面还要讨论)，还是能够识别内含子与外显子的边界的。尽管如此，依靠计算机的方法并不能做到完全准确。大约有 3/4 的基因能够通过这种方法被预测，还有很多则不能预测出来，即使在所识别的预测基因中，一些短的外显子，尤其是非编码外显子常常被错误地鉴定出来。

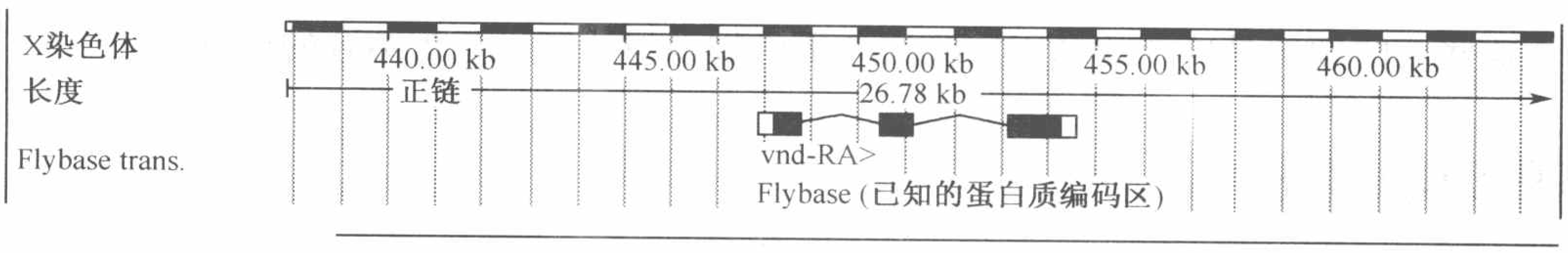


图 20-1 果蝇 *vnd* 位点的结构。图示为 X 染色体上一段长约 25kb，含有 *vnd* 基因的区域。*Vnd* 基因转录物含有三个外显子和两个内含子。其中 5′（左）和 3′（右）端空白框表示非编码序列，即不参与最终蛋白质的编码。FlyBase 是一个用于果蝇基因组分析的标准数据库。

用全基因组覆瓦式芯片分析转录组

在完成一个生物的全基因序列组装后，它可以用来全面分析特定细胞或组织中表达的所有蛋白质编码基因或者非编码序列（如内含子与 miRNA 基因）。全基因组的代表性转录物被称为转录组（transcriptome)。将人工合成的长 50 个核苷酸的单链 DNA 片段（寡核苷酸）点到玻璃片或硅片上。大约每隔 100～150 bp 的基因组序列设一个这样的寡核苷酸，按基因组序列依次排列，即是“覆瓦式芯片”（意为像屋顶的瓦片那样覆盖整个基因组的芯片——译者注)。全基因组的覆瓦式芯片发展非常迅速。现在，该技术已经发展到可以在 1cm^2 大的玻璃片或者硅片上制备覆盖全基因组的完整的芯片。例如，1 百万个 50mer 的寡核苷酸可以点到同一张芯片上，可覆盖果蝇的整个基因组。我们在下面将讨论到，芯片上每个点（即每一段核苷酸序列）只能通过安装在显微镜上的微型感应器才能检测其杂交信号。

为了获得转录组信息，将覆瓦式芯片与荧光标记的 RNA（或 cDNA）探针进行杂

交。这些探针可以来自特定类型的细胞，如海鞘幼虫尾部肌肉细胞或在特殊培养基里生长的酵母细胞。最终显示的结果是得到一系列杂交信号，对应于整个基因组中所有预测的蛋白质编码序列（图 20-2）。

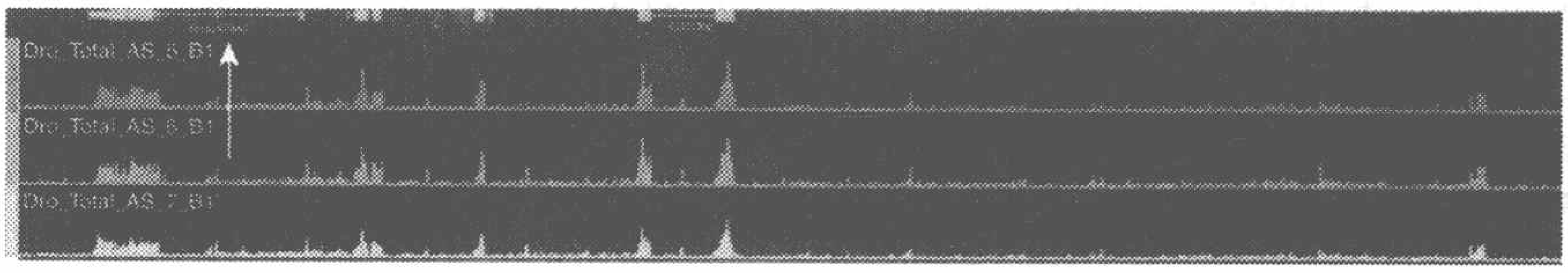


图 20-2　全基因组覆瓦式芯片。果蝇 3 号染色体上一个长 50kb 的区段，含 4 个基因。图上部所示为各个转录本的外显子和内含子的结构。白色箭头所指为一很大的内含子区，可能含有一个小外显子（micro-exon）。总 RNA 取自不同发育阶段的胚胎，与全基因组覆瓦式芯片杂交。此芯片由基因组序列上每隔 35bp 一个的 25bp 的寡核苷酸组成。杂交信号的强和弱对应于外显子和内含子。图中红色表示早期胚胎、绿色和蓝色表示发育较老的胚胎。（经许可引自 Manak *et al*. 2006. *Nat. Genet*. 38：1151-1158，Fig. 5. © Macmillan.）

全基因组覆瓦式芯片提供了各个转录单元中的内含子与外显子结构的直接信息。由于内含子转录物的不稳定性，一旦从初级转录物中出来后便会迅速地降解。这些实验使用的都是总 RNA（total RNA），但是由于外显子序列较内含子序列稳定，经过标记和杂交，外显子序列显示的信号要比内含子序列的信号强得多。

全基因组覆瓦式芯片的另外一个优点是能用于检测非编码基因，如 miRNA 基因。miRNA 是由一些大片段的 RNA 前体（pri-RNA）加工而来。其转录单元长度为 1～10kb（见第 18 章）。通过与覆瓦式芯片杂交能很容易地检测这些 pri-RNA。在某些情况下，miRNA 基因也含有内含子。只有在这些内含子被剪切之后，该基因才能生成成熟的 miRNA。此外还有一些其他的非编码转录物同样可被检测，其中包括蛋白质编码基因的内含子中的“反义”RNA（antisense RNA）。这些 RNA 可能在调控蛋白质编码基因的表达和功能方面发挥作用。

用覆瓦式芯片得到了一些令人震惊的结果：在一个典型的基因组中，大约有 1/3 的序列是转录的，尽管只有其中一小部分是蛋白质编码序列。例如人类的蛋白质编码序列只占基因组的 5%。这说明大多数转录物来自基因组中大量的内含子 DNA 序列。许多基因在离基因编码序列很远处（1 Mb 或更远）有 5′端的非编码外显子。这些序列，有时候能相隔一百万个碱基。在某些情况下，这些内含子区域能够产生 miRNA 和一些其他类型的非编码 RNA。

专门的比对工具可用来鉴定 DNA 调控序列

基因组技术在鉴别基因和分析转录单元的结构方面十分有效。基因鉴定后，生物信息学方法能够预测该基因编码蛋白质的可能功能和结构。例如，是否含有已知的结构域（domain）或基序（motif），或者是否与已知蛋白质具有相似的特性。在这些方法中，基本局部比对搜索工具（BLAST，the Basic Local Alignment Search Tool），即 BLAST 算法提供了对蛋白质或者核苷酸序列搜索、比较和比对的一种强大的方法。通过

BLAST 搜索，一段外显子序列可以被快速地同庞大的蛋白质编码数据库进行比对（图 20-3）。显著比对上的蛋白质编码序列的功能，如 DNA 结合蛋白质、复制因子或细胞膜受体等，可以用于快速预测该基因的可能活性和蛋白质产物。此外，简单的 BLAST 搜索还能识别产生 miRNA 的非编码转录物（见第 18 章）。

```
Size=6 Zinc finger protein SLUG (Neural crest
transcription factor Slug) (Snail homolog 2)
Length = 268

Score = 217 bits (553), Expect = 1e    -56
Identities = 98/144 (68%), Positives = 113/144 (78%), Gaps = 4/144
(2%), Frame =   -1

Query: 264 LPCKCPICGKAFSRPWLLQGHIRTHTGEKPFQCPDCPRSFADRSNLRAHQQTHVDVKKYA
           LPC C ICGKAFSRPWLLQGHIRTHTGEKPF CP C R+FADRSNLRAH QTH DVKKY
Sbjct: 183 LPCVCKICGKAFSRPWLLQGHIRTHTGEKPFSCPHCNRAFADRSNLRAHLQTHSDVKKYQ

Query: 84  CQVCHKSFSRMSLLNKHSSSNCTI 13
           C+ C K+FSRMSLL+KH  S C +
Sbjct: 243 CKNCSKTFSRMSLLHKHEESGCCV 266
```

图 20-3 编码一个调控蛋白质的外显子的 BLAST 分析。 查询（query）序列所对应的是用来分析的未知基因组区段，并在庞大的蛋白质编码序列数据库进行搜索。目标（Sbjct）序列表示同未知序列相似度最高的序列。在本例中，目标序列对应的是调控蛋白质 SLUG。

在第 17 章我们已经了解到，在基因编码序列之外还存在着许多 DNA 调控序列，这些序列调控着相关基因在机体内何时何处表达或关闭，对于这些序列的识别和分析更具挑战性。的确如此，这些序列被称为基因组中的“暗物质”（dark matter）。如今，只有全基因组学的方法能够用于鉴定这一类重要的 DNA 序列信息。

BLAST 搜索是为识别蛋白质编码序列而建立的，稍作变化也可用于鉴定脊椎动物的一部分调控序列。细胞特异性增强子含有一个或多个序列特异性 DNA 结合蛋白质的位点簇（见第 17 章）。在某些情况下，这些位点簇足以用于鉴定一些可以比对的短 DNA 序列。计算机程序 VISTA 能以 10～20bp 的窗口对不同基因组进行比对，因此可以鉴定仅长 50～75bp 的不完全保守的非编码序列（图 20-4）。河豚鱼和小鼠具有 1 万个相似的非编码短序列，可以相信其中很多都有相应的组织特异性增强子。然而，这两种动物，尤其是小鼠，很可能有 10 万个增强子。由此而知，这些简单的序列比对方法并不能完全识别数量庞大的调控序列。

组织特异性增强子还可以通过扫描基因组中的已知调控蛋白质的潜在结合位点（potential binding site）来鉴定。例如编码一个细胞黏附分子的 α-环素（α-catenin）基因在许多不同的组织中都有表达，但在心脏前体细胞（心肌细胞）中的表达尤高。因此，就有可能通过对 α-环素基因的两侧和内含子序列的比对分析，寻找已知心脏调控蛋白质（如 MEF2C、GATA -4 和 E47/HAND）的结合位点，来鉴定可能存在的心脏特异性增强子（图 20-5）。这些调控蛋白质每一个都能识别一些长 6～8 bp 的 DNA 基序。这些结合位点可以用位权矩阵（PWM，position-weighted matrix）来描述，能够用不同的计算机和实验的方法，如 SELEX（in vitro selection，“体外选择”）来决定（详见第 21 章）。应用位权矩

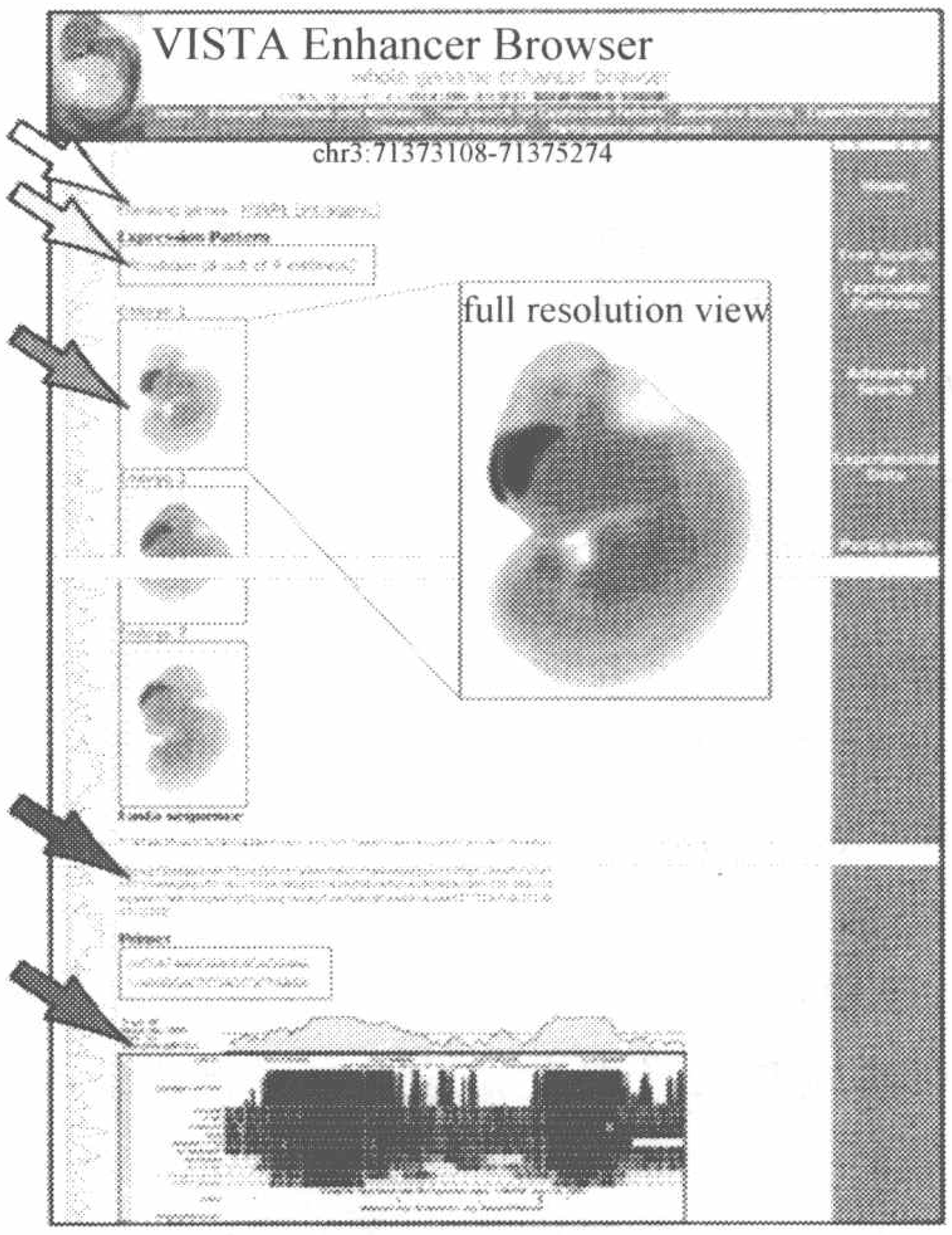


图 20-4 调控序列的 VISTA 比对。图下部的红色箭头显示小鼠基因组中两个高度保守蛋白质编码基因之间非编码区的保守序列。这一非编码序列接到 *lacZ* 报道基因（reporter gene）上，并在转基因小鼠的胚胎中表达。图中绿色箭头和方框所显示的是非编码保守序列所指导的表达方式。该序列作为增强子，其功能是指导 *lacZ* 报道基因在鼠胚胎的前脑发育过程中的表达。（经许可，复制自 Visel A. *et al*. 2007. *Nucleic Acids Res*. 35：D88－D92，Fig. 1© Oxford University Press.）

阵来搜索 α-环素基因的位点，就能鉴定 MEF2C、GATA -4 和 E47/HAND 的结合位点簇。实验方法也证明这些结合位点簇位于基因的 5′外侧，真正具有增强子的功能。

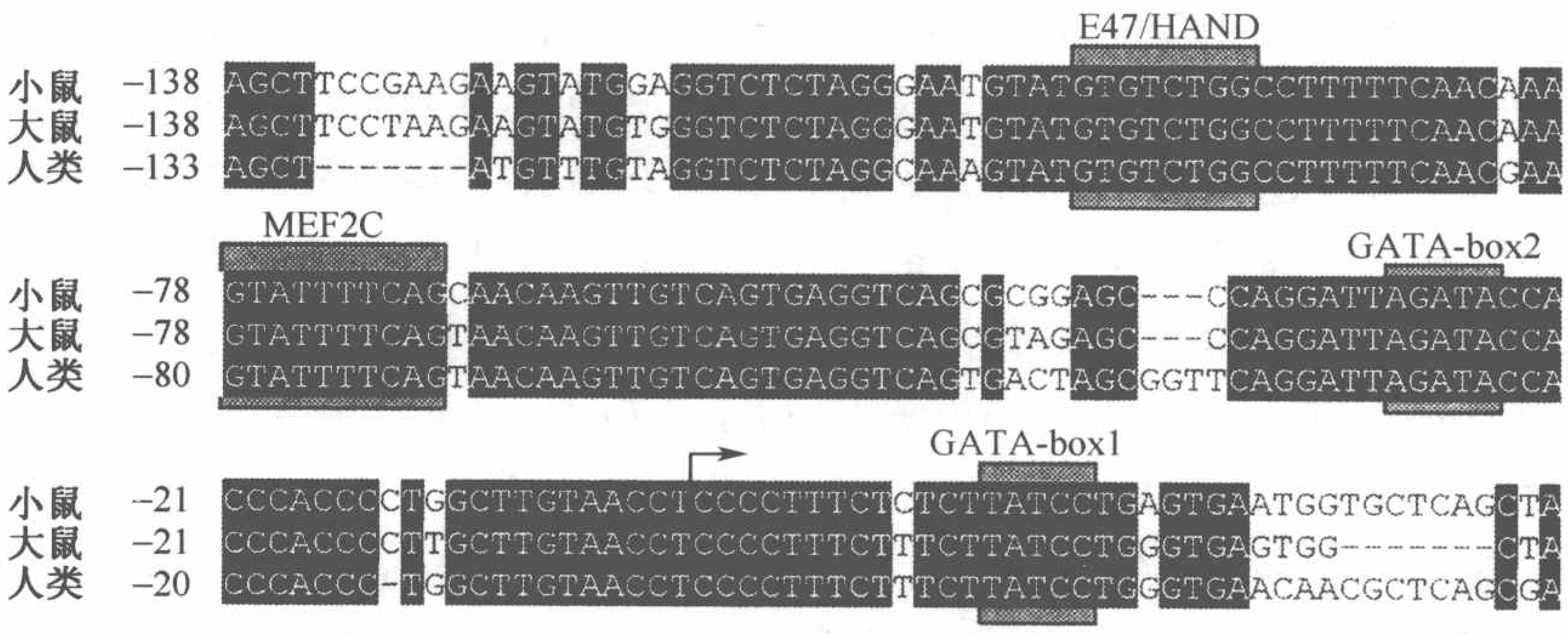


图 20-5 心脏特异性增强子的 in silico 鉴定。α 连环蛋白质基因 5′侧区域的一段长约 140bp 的序列，它在小鼠、大鼠和人类基因组中具有保守性。该保守序列含有三个关键的心脏分化调控蛋白质（E47/HAND、MEF2C 和 GATA）结合位点。实验已经证明，该序列在小鼠中是一个真正的心脏特异性增强子。理论上说，该序列可以通过 VISTA 比对（见图 20-4）或者心脏调控蛋白质聚类分析而被鉴定。（经许可，复制自 Vanpoucke G.，*et al*. 2004. *Nucleic Acids Res*. 32：4155－4165，图 1 局部.）

按理说，通过在小鼠或人类全基因组搜索 MEF2C 和 E47/HAND 结合位点簇，其他心脏特异性增强子应该也能被鉴定出来。这一方法已经被应用于增强子的鉴定，但是效果甚微（见框 20-1）。就算是最好的结果，在受某些调控因子结合调控的所有增强子中，也仅有 10%～20%的增强子能被此方法识别。此外，在这些识别的候选增强子中，许多是假阳性，并不具有增强子的功能。当该方法应用于特定的候选基因时却能够获得较好的成功，如上所述的 α-环素基因。

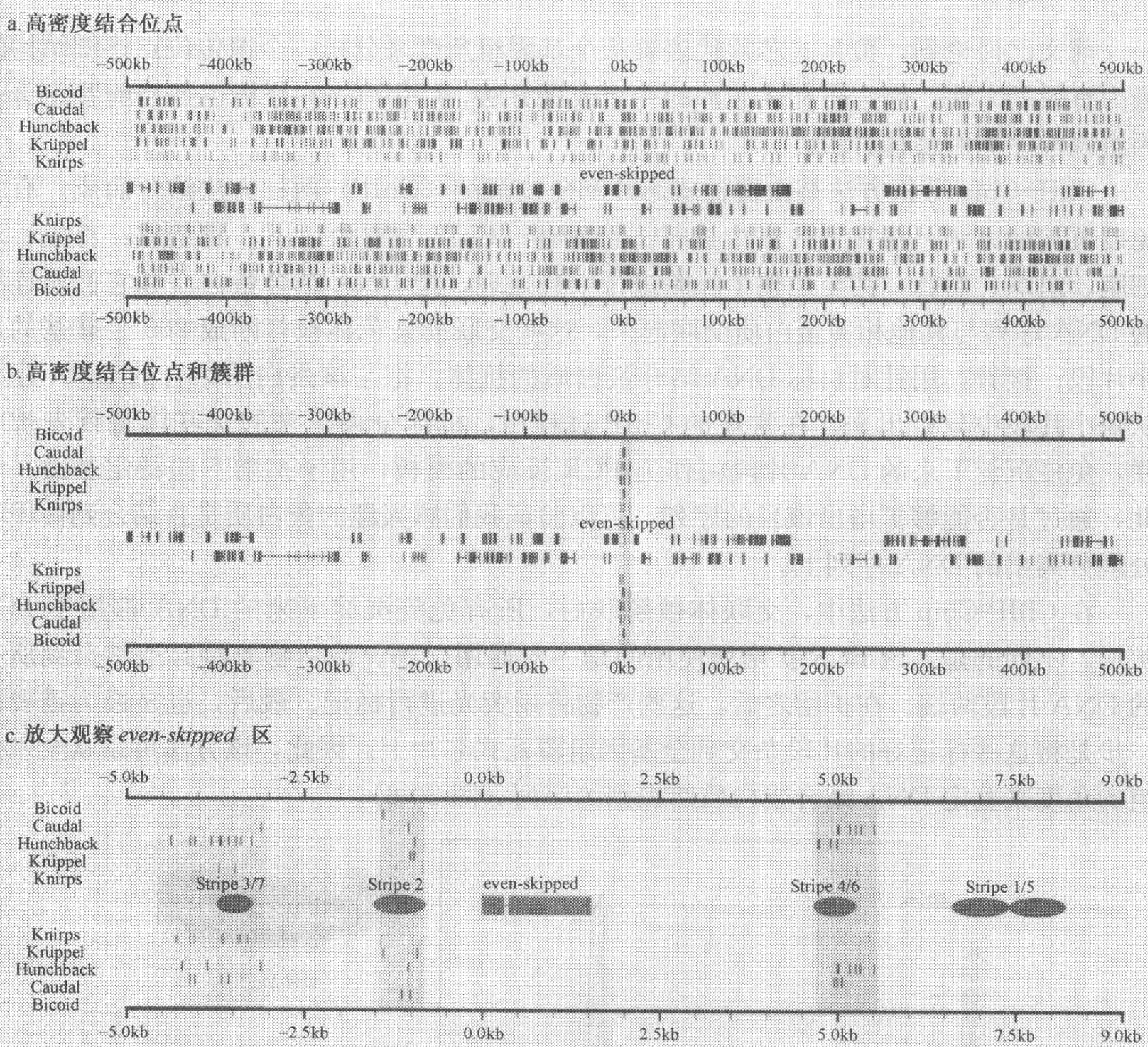
框 20-1 识别复杂增强子的生物信息学方法

多种计算机程序，称为“全基因组组装”已被开发出来，并用来从已有的基因组序列中鉴定调控 DNA。这些程序利用 DNA 调控序列的一个重要特征——调控区常有密集的 DNA 结合位点簇。例如 eve stripe-2 增强子长 500bp，含有 4 个调控蛋白质的 12 个相隔的结合位点。这 4 个调控蛋白质分别为 Bicoid、Hunchback、Krüppel 和 Knirps（见图 19-21）。因此，在该增强子中，大约每 50bp 就含有一个以上的结合位点。这种密集分布的结合位点，在果蝇胚胎早期发育过程中调控基因局部表达方式的增强子中尤为典型。

正如本章中所讨论的那样，许多调控蛋白质已经被证明参与调控果蝇胚胎成对条纹（pair-rule stripe）基因的表达。这些调控蛋白质包括 Bicoid、Hunchback、Krüppel、Giant 和 Knirps。不幸的是，到目前为止，已鉴定的 Giant 结合位点还不足以用来确定该蛋白质可能识别的调控序列范围。相反，其他 4 个调控蛋白质以及另外一个同源异型结构域蛋白质 Caudal 都有足够的 DNA 结合域的信息。Caudal 是一个转录激活子，并在胚胎的后半部分呈现出广泛梯度表达模式。

Bicoid、Hunchback、Krüppel、Giant 和 Knirps 以单体的形式结合到 DNA 上。它们所识别的序列相对简单，但在果蝇基因组中拷贝数目极大。例如，Bicoid 识别的序列是一段由 ATTA 核心基序和一些侧翼 C/G 残基所组成的简单序列。在果蝇基因组中，平均约每 1kb 就含有一个 Bicoid 的潜在结合位点。因此，如果只用 Bicoid 结合位点来识别基因组上片段分布的增强子将是徒劳的。因为在其基因组中，一共有 10 万个以上的该位点（几乎每个基因有 10 个）。但是，如结合 Bicoid 蛋白质和与其共同作用调控蛋白质的结合位点簇，将能够去除众多的假结合位点（或者称为噪音）。

拿一段长 1Mb，且含有 eve 位点的 DNA 序列来说（框 20-1 图 1）。该 DNA 含有成千上万个 Bicoid、Hunchback、Krüppel、Giant 和 Knirps 结合位点（框 20-1 图 1a）。但是，在该序列中，以 700bp 或更小的窗口，只有三簇含有 13 个以上的结合位点，即其密度为约每 50bp 一个结合位点（框 20-1 图 1b）。值得注意的是，这三簇同 eve 基因的 5′和 3′端调控区域相一致。其中一簇与 eve stripe—3/7 增强子对应，第二簇与 eve stripe-2 增强子对应，第三簇位于 eve 基因的 3′调控区域，并与 eve stripe-4/6 增强子对应（框 20-1 图 1）。



框 20-1 图 1　结合位点聚类分析鉴定 eve 基因 stripe 增强子。(a) 一段长 1Mb 包含 *eve-skipped* 基因座序列，分布着 Bicoid、Caudal、Hunchback、Krüppel 和 Knirps 结合位点。*Even-skipped* 基因座位于该区域的中间，侧翼序列为相邻基因的内含子-外显子序列。(b) 在这段 1Mb 序列中，高密度结合位点簇仅出现在 *eve* 基因的附近。(c) 共有 3 处高密度结合位点簇与 *eve* 基因相对应。这 3 个结合位点簇的位置与 stripe-3/7 、stripe-2 和 stripe-4/6 增强子的位置分布一致。（经允许，摘自 Berman P. *et al*. 2002. *Proc. Natl. Acad. Sci.* 99：757-762，图 1，p159. 版权所有，美国国家科学院.）

已经证明，DNA 结合位点聚类分析是果蝇基因组增强子鉴定过程中非常有用的工具。但是，目前的计算机程序并非具有 100%的准确性。即使是最好的预测软件，也大约只能够成功预测 1/3 的真正增强子。设想，假如在结合位点的组合基础上加入空间限制，将会比以往简单的位点簇的方法获得更高的命中率。例如，在第 17 章我们已看到，干扰素增强子体包含有固定空间结构的结合位点，其中包括相连位点之间的螺旋相位。

增强子识别的最佳方法——ChIP-Chip 分析

前文已讨论到，覆瓦式芯片代表着从全基因组角度来分析一个遗传位点详细结构的最为有效的方法。作为覆瓦式芯片的一个改进方法，ChIP-Chip 分析也是当前鉴定全基因组调控序列的最佳方法。

ChIP-Chip 分析方法是由覆瓦式芯片和免疫沉淀（ChIP）两种方法结合而成。有关免疫沉淀的方法，将在下一章中进行详细阐述。ChIP 主要由以下步骤组成：首先，对细胞、组织、器官，甚至是整个胚胎进行甲醛处理，把 DNA 结合蛋白质和它们所连接的 DNA 序列与其他相关蛋白质交联起来，这些交联的染色体被打断成 200 个碱基的短小片段；接着，用针对目标 DNA 结合蛋白质的抗体，把与该蛋白质结合的 DNA 序列从短小片段中分析出来。在常规的 ChIP 过程中，抗体分离出来的交联体将首先被解联，免疫沉淀下来的 DNA 片段将作为 PCR 反应的模板，用于扩增一些特定基因。因此，通过是否能够扩增出该目的序列，可以验证我们感兴趣的蛋白质是否结合到由甲醛处理分离出的 DNA 序列上。

在 ChIP-Chip 方法中，交联体被解联后，所有免疫沉淀下来的 DNA 都用于 PCR 扩增。不同的是，该 PCR 扩增所使用的是一对通用引物。该引物非特异地结合到所有的 DNA 片段两端。在扩增之后，这些产物将用荧光进行标记。最后，也是最为重要的一步是将这些标记好的片段杂交到全基因组覆瓦式芯片上。因此，该方法可以从全基因组的角度来鉴定 DNA 结合蛋白质所识别的序列（图 20-6）。

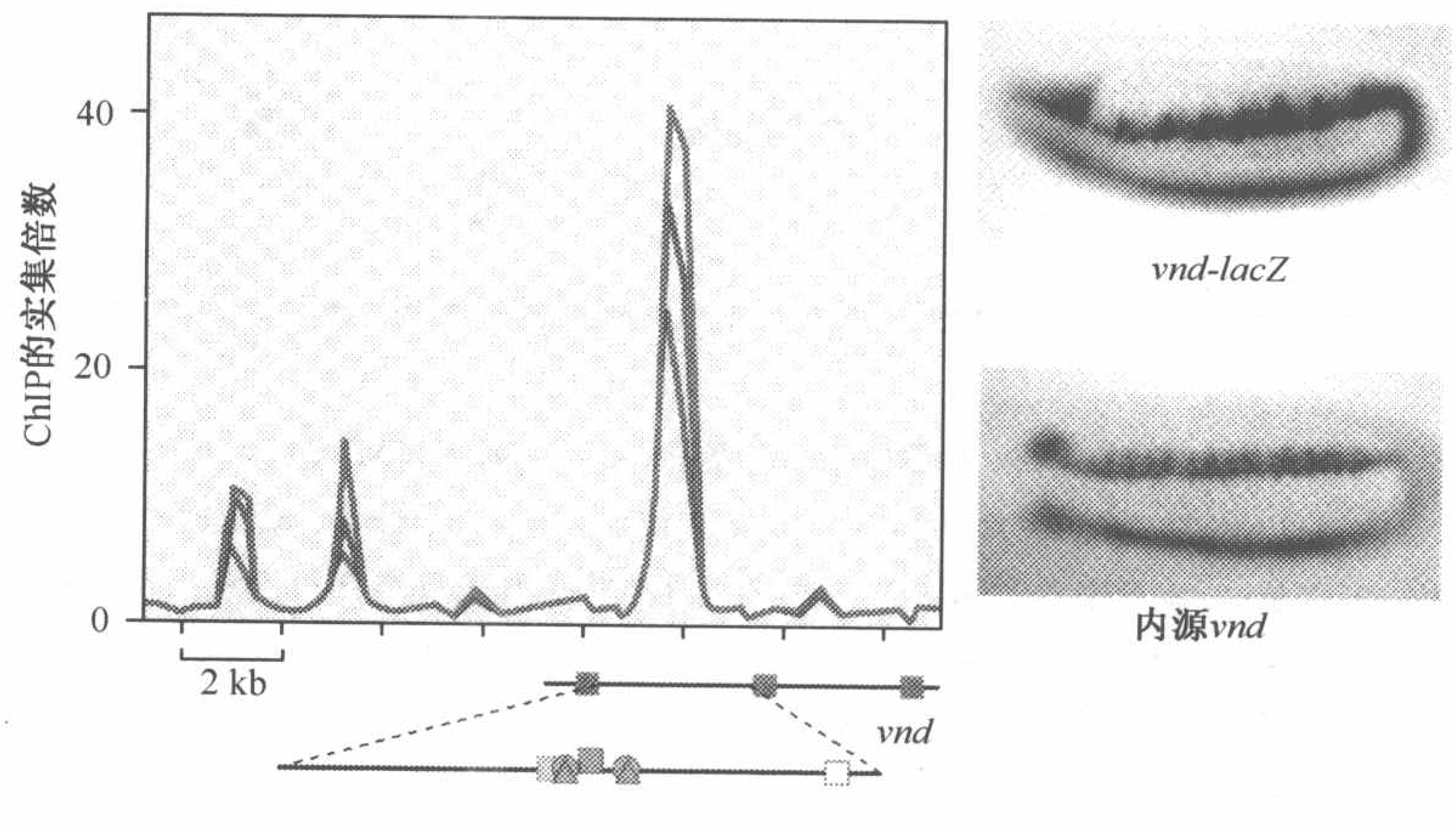


图 20-6　用 ChIP-chip 方法鉴定 *vnd* 位点 DNA 调控序列。与调控蛋白质 Dorsal（红色）、Twist（蓝色）以及 Snail（绿色）相互作用的 DNA 片段，分别杂交到果蝇全基因组覆瓦式芯片上。图中，左边所示的是覆瓦式芯片上包含 *vnd* 位点的区域。在 *vnd* 转录物内含子 1 中，包含有一个 Dorsal、Twist 和 Snail 结合位点的大峰。在其 5′外端还包含两个小峰。通过 *lacZ* 报道基因以及在转基因胚胎内的表达证明，包含有任何一个峰所在序列的基因组 DNA 都可以作为一个真正的增强子。图右方所示为由内含子增强子序列引导 lacZ 在胚胎中表达的图谱。图右下显示的是内源 *vnd* mRNA 在胚胎中的分布。（经允许，摘自 Zeitlinger J. *et al*. 2007. *Genes Dev*. 21：385-390，图 3b. 冷泉港实验室出版社出版.）

ChIP-Chip 具有强大的功能，它已经成功应用于定位许多调控蛋白质在全基因组上的分布。其中的一个例子便是该技术在鉴别序列特异性转录因子 Nanog、Sox2 和 Oct4 上的应用（图 20-7）。这些调控蛋白质部分地决定着人类胚胎干细胞的特性，如自我繁殖能力和能够分化成各种不同细胞类型的能力（详细的介绍请参考框 17-6 和框 19-6）。

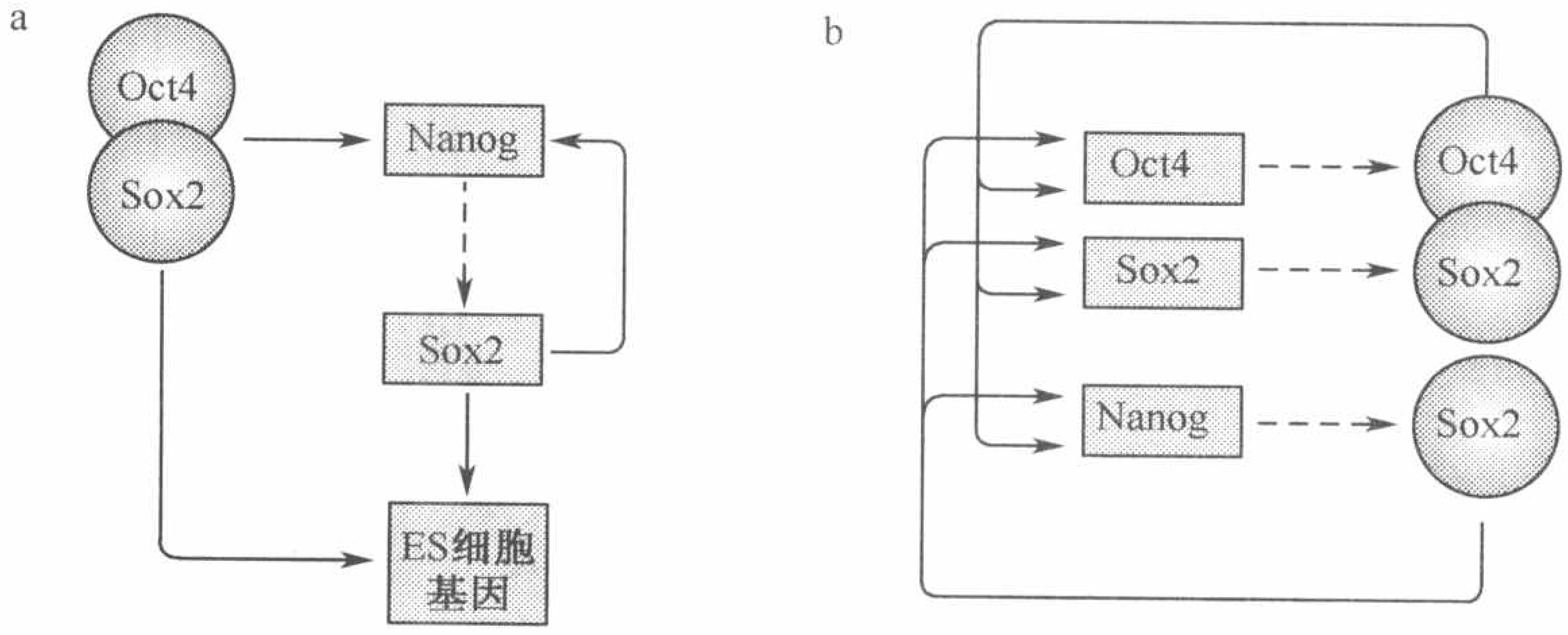


图 20-7 决定干细胞特性的蛋白质。调控蛋白质 Oct4 和 Sox2 形成复合物，激活 Nanog 的表达。Nanog 的表达也受 Sox2 的独立调控。Oct4、Sox2 和 Nanog 是维持胚胎干细胞状态的关键蛋白质。

针对 Nanog、Sox2 和 Oct4 的特异抗体已经应用于鉴定这些蛋白质在干细胞中的结合位点（参考图 20-8）。结果显示，这 3 个调控蛋白质能够共同调控细胞中 100 个以上的可能靶增强子。在这些靶增强子中，一部分是同已知的发育调控基因密切相关，例如 *Hoxbl* 基因，该基因与果蝇中的同源异型基因功能相似（见第 19 章）。

覆瓦式芯片和 ChIP-chip 分析方法的结合，在全面鉴定特定生物过程中所涉及的 DNA 调控序列方面显示出强大的优势。在此，我们以 α 环素基因和一些在发育中的心脏中特异表达的基因为例来做说明。从心脏细胞中提取总 RNA 后和全基因组覆瓦式芯片杂交。可以鉴定出在不同发育阶段的心脏前体细胞中特异表达的基因，包括编码基因和非编码基因。然后，用 ChIP-chip 分析方法鉴定 MEF2C、E47/HAND 和 GATA4 蛋白质在体内的结合位点，这几个蛋白质是前文描述的心脏组织特异表达基因的关键调控因子。结合 ChIP 的显著信号，以及覆瓦式芯片的基因表达分析，能够很好地鉴定大部分或者全部的心脏细胞特异的增强子。

许多的动物间有非常相近的基因群

下面我们将讨论比较基因组分析的一个例子，主要是动物基因组比较的结果。迄今为止，大约有 100 种不同动物的基因组被完全测序并完成拼接，但是其中大部分序列都局限于少数的以人类为中心的动物类型，以及主要的模式动物类，如果蝇（*Drosophila melanogaster*）以及线虫（*Caenorhabditis elegans*）。因此，为了研究人类基因组特点，许多灵长类动物基因组（例如黑猩猩、猕猴等）都已进行了完全测序。果蝇的 12 个亚种已经测序完毕，并用来研究其不同种属的多样性。现在，仅有 1/3 的动物门中至少有一个代表动物被测序并完成组装。

迄今为止，基于比较基因组序列分析的发现中，最让人吃惊的是在数量庞大的动物种群分支中，从海葵到人类都拥有一些相似的保守基因。无脊椎动物，例如海葵、蠕虫

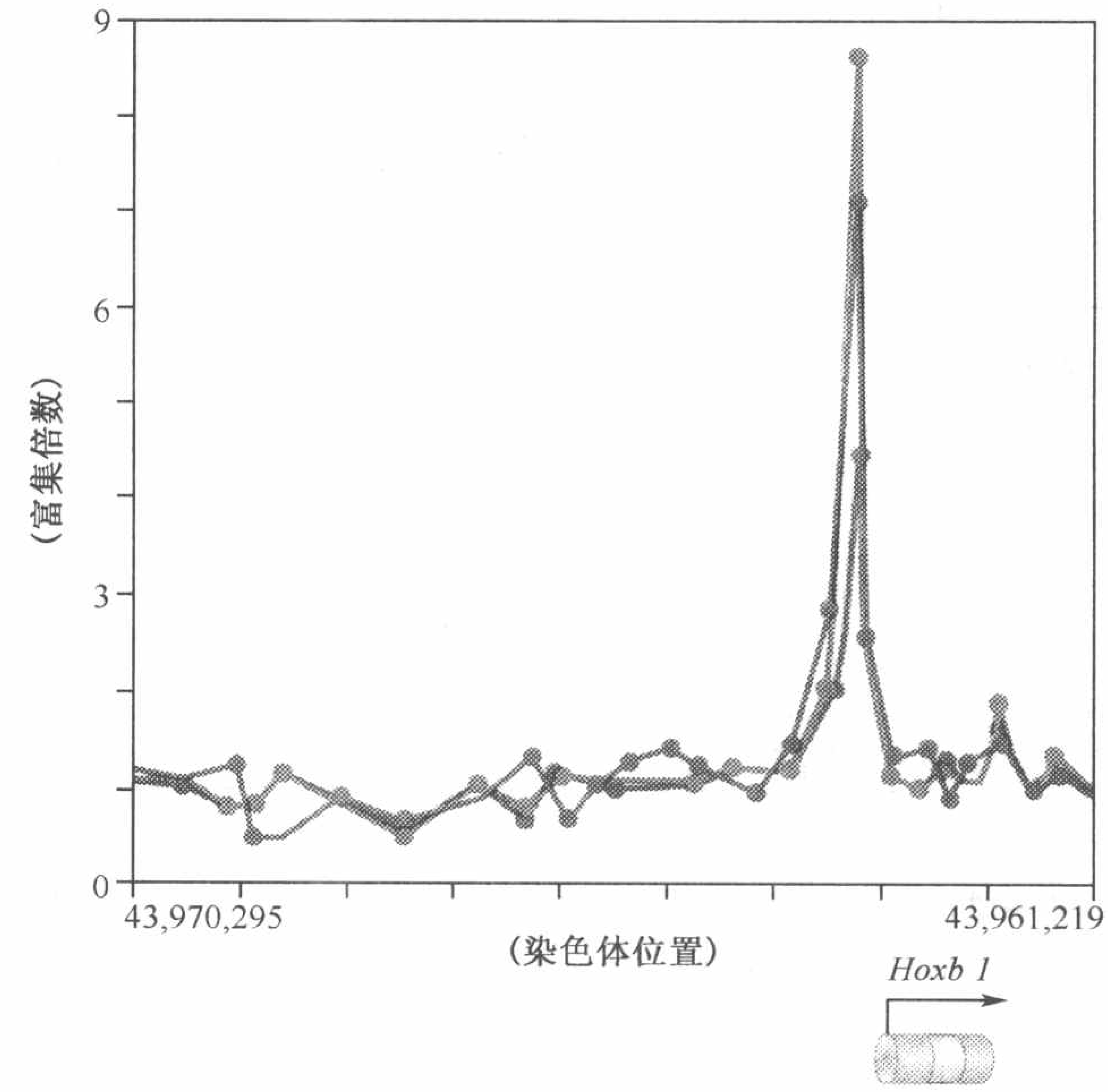


图 20-8　用 ChIP-chip 鉴定被干细胞因子调控的增强子。与 Nanog（绿色）、Sox2（红色）和 Oct4（蓝色）交联的 DNA 片段杂交到人类全基因组覆瓦式芯片上。这 3 个蛋白质都集聚在与 *Hoxb1* 基因相关的 5′端侧翼序列上。（经允许，摘自 Boyer L. A. *et al*. 2005. *Cell* 122：947-956，图 2b。Elsevier.）

以及昆虫，都含有大约 15 000 个蛋白质编码基因。脊椎动物含有的蛋白质编码基因的数量稍微多一些，大约有 25 000 个。但是，那些多出来的基因并不是在脊椎动物中特有的全新基因，而是来源于那些存在于无脊椎动物基因组中的“老基因”的复制。例如，在无脊椎动物中纤维细胞生长因子（FGF）基因仅有少量拷贝，而在典型的脊椎动物基因组中，该基因含有超过 20 个拷贝。

对动物差异化之属性所必需的基因群的了解，原生单细胞真核生物 Monosiga 的基因组测序拼接提供了很多信息。该生物是最接近现有动物种类的物种。但是，它缺乏许多与动物发育相关的编码基因，包括信号分子编码基因如 Wingless、转化生长因子 β（TGF-β）即 Dpp、Hedgehog 以及 Notch（见第 19 章）。此外，它还缺乏在动物胚胎发育过程中操纵不同基因活性的关键调控基因，例如 Hox 基因以及 Hox 簇。因此，从这种简单的真核生物到现代生物的进化过程中，需要大量的在十亿年前的远古深海生物中没有的新型基因的出现。

许多动物中都含有不规则基因

尽管所有动物都含有被称为“工具箱”的保守基因群，这些基因是所有动物生长发育过程中必需的。此外，每一种动物的基因组还含有决定自己特性的基因，有时候，这类基因能够带来意想不到的作用。例如海鞘，这种动物含有编码纤维素合酶的基因（图 20-9）。在植物中，这种酶应用于编码树木的主要组成部分——纤维素。事实上，纤维素

在动物中几乎是不存在的。那么，该基因对于海鞘有什么作用呢？成年海鞘是不动的，它们在潮汐的作用下过滤海水（见第 19 章）。海鞘具有一层富有弹力的保护性外鞘。这种外鞘的主要组成部分是动物纤维素，一种同植物纤维素相类似的物质。但是，通过海鞘基因组测序拼装的初步分析，我们无法得知它们是拥有自己的纤维素合成系统，还是借由别的生物来生产这种外鞘。因为的确是有一些生物借由简单的共生体来完成罕见的功能。例如在白蚁和一种吃木头的蟑螂的后肠中就含有共生的细菌来帮助它们消化木屑。

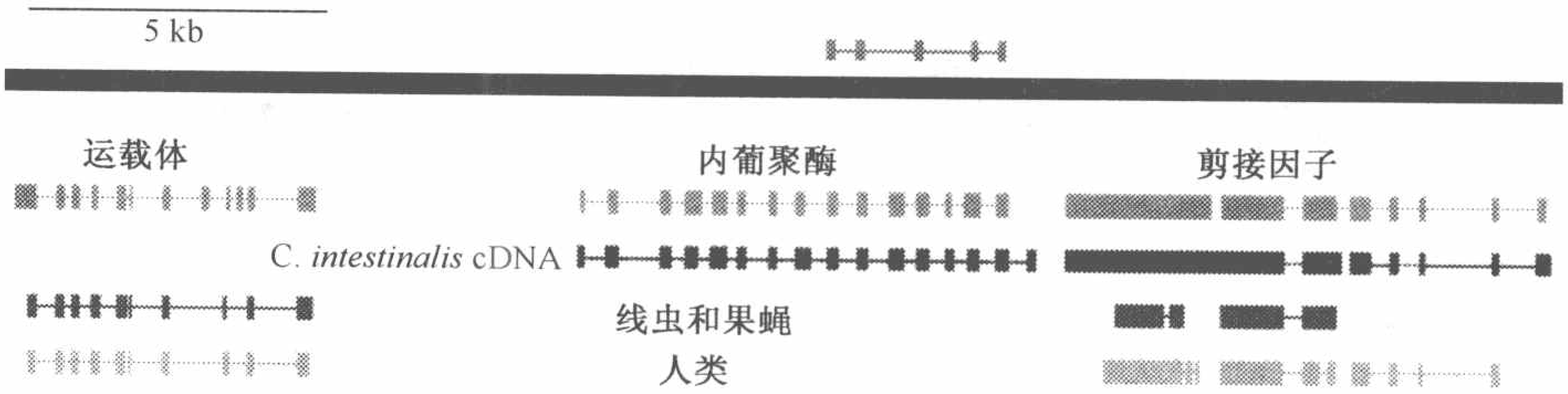


图 20-9 玻璃海鞘（*Ciona*）基因组中的一个植物基因同其他动物序列的比对。图示为 *Ciona* 基因组中长 20kb 的区域。该区域含有一个内葡聚酶基因。该基因编码一个与纤维素降解和合成相关的关键酶。纤维素是植物细胞壁的主要组成成分。图中红色矩形条表示拟南芥 *Kerrigan*-1 基因。通过基因搜索软件，在 *Ciona* 基因中一共找到 15 个可能的外显子，如图中绿色所示。实际上，在该基因的 5′端还存在一个外显子（黑色框），而计算机软件却未能将其找到。与此类似，计算机还预测到一个编码 RNA 剪接因子的侧翼基因，并在预测到其大片段编码序列中存在着一个内含子。而事实上，cDNA 序列说明该小内含子并不存在。在 5′最外端的外显子，预测结果同实际结果也不完全一致。在线虫、果蝇和人类之间，其侧翼基因是保守的，而内葡聚酶基因唯一存在于 *Ciona* 中。*Ciona* 含有一个纤维性外鞘。需要提醒的是，在所有不同动物的基因组中，其侧翼基因的外显子-内含子结构并不完全相同。（经允许，摘自 Dehal *et al*. 2002. *Science* 298：2157-2167，图 7。AAAS.）

对于海胆基因组的研究带来了另外一个惊喜：它含有两个在人类和其他脊椎动物中免疫球蛋白基因重排所必需的基因：*RAG1* 和 *RAG2*（见第 11 章）。脊椎动物的一个重要特点是拥有免疫系统，对感染以及外界伤害作出免疫反应。这种免疫反应包括产生抗体用于对外来抗原的高精度、高特异性的识别以及结合。无脊椎动物也有其先天的免疫系统，但是它们却不能产生适应性免疫反应。通过对海胆基因组的初步分析，人们推测现代脊椎动物的祖先可能和一种含有 *RAG1* 和 *RGA2* 基因的细菌相结合。但是通过对海胆基因的分析表明，这种猜想是错误的。相反，*RAG* 基因为另外一种远古动物——现棘皮动物（例如海胆）和脊索动物（例如脊椎动物）的祖先所必需的（见图 20-10）。似乎是在进化的过程中，这个假设祖先的部分后代逐渐失去了 *RAG* 基因，例如海鞘。

进化中的同线性

比较基因组分析所发现的最令人震撼的结果是高度同线性（synteny）的。同线性是指遗传连锁群在遗传距离很远的动物间的保守性。在小鼠与人类之间，它们的同线性程度很高。在许多情况下，同线性甚至能够延伸至河豚鱼。在 4 亿多年前，河豚鱼与许多哺乳动物属于共同的祖先。更为甚者，人类与某些简单的无脊椎动物具有一些共同的保守同线性。海葵便是其中典型的例子。7 亿年前，海葵与人类属于共同的祖先。这时

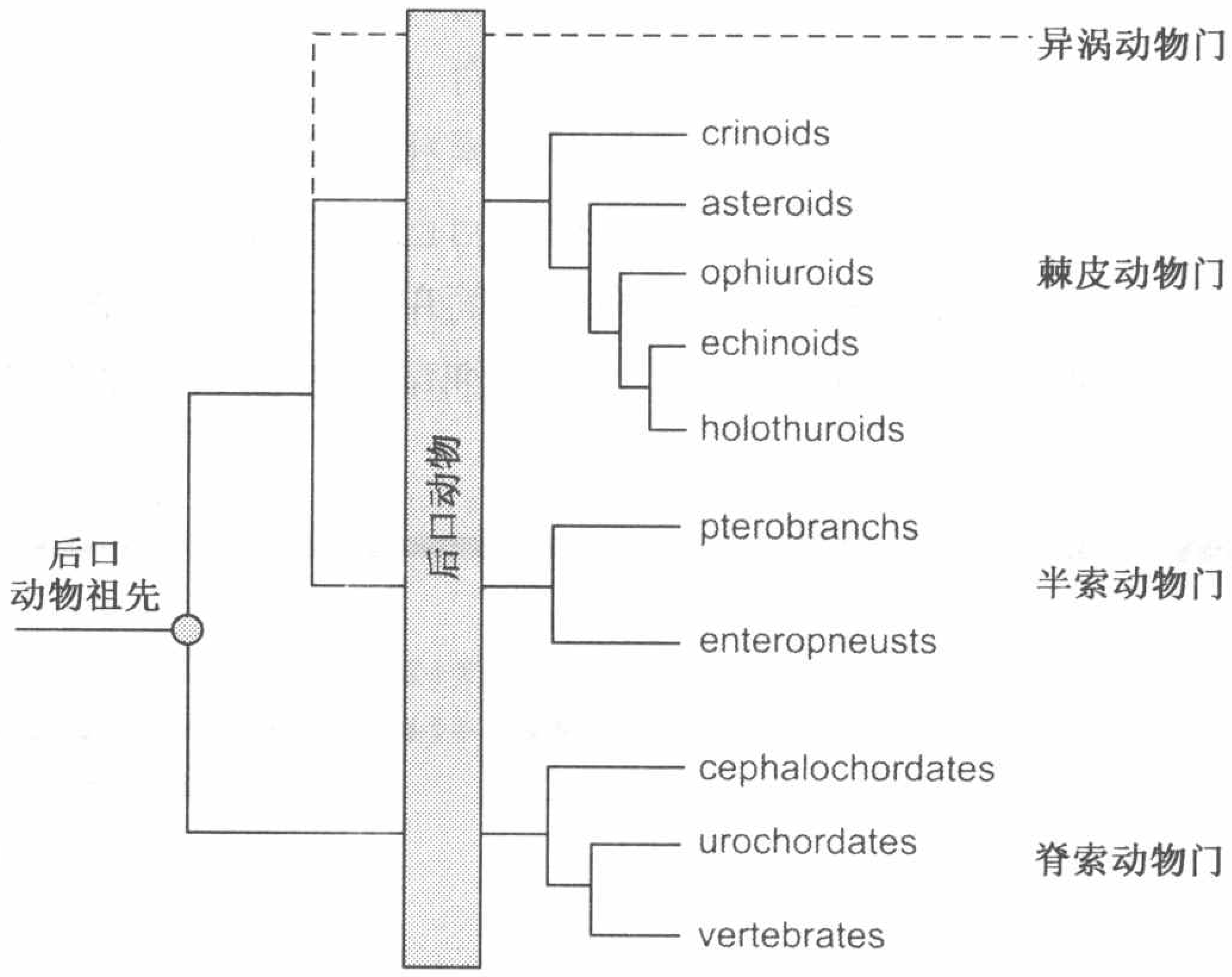


图 20-10 后口动物进化史。后口动物一共有 4 个门：异涡动物门、棘皮动物门、半索动物门和脊索动物门。棘皮动物门一共有 5 个种，而脊索动物门有 3 个种。值得注意的是，与脊椎动物最为接近的是尾索动物。其中海鞘就是尾索动物的一种（参考框 19-3）（经允许，摘自 Gerhat J. 2006. J. Cell Physiol. 209：677-685.）

间正处于寒武纪辐射，大量现代动物种类形成期之前（图 20-11）。

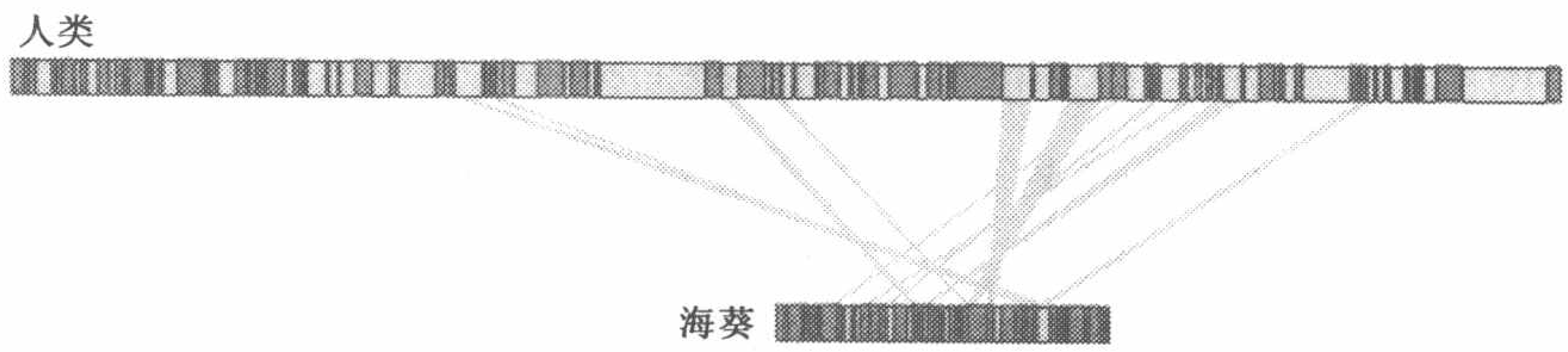


图 20-11 海葵与人类遗传连锁的保守性。上图显示的是人类第 10 号染色体 q24 上 10Mb 的区域。图中线条所表示的是在该区域中的 11 个基因与海葵染色体中 1Mb 对应区域的比对结果。这 11 个基因都位于两个物种的染色体中，所不同的是，在 7 亿年的进化过程中，这 11 个基因在人类和海葵基因组中的排列顺序发生了变化。

对原核生物来讲，遗传连锁至关重要，连锁的基因往往受一个共同的操纵子所调控（参考第 16 章）。但在后生动物中，此类连锁几乎遗失殆尽。这也就是说，相邻的基因比不相邻的基因不一定有更紧密的共表达（如在血液细胞中）。早期比较基因组分析似乎已经确定，遗传连锁对基因的调控并没有多大的影响。例如，在哺乳动物基因组之间（如小鼠与人类）或是无脊椎动物基因组之间（如线虫与果蝇），相关基因之间的排列并没有明显同线性关系。然而，越来越多的证据表明，线虫同果蝇基因组是高度“派生的”。也就是说，它们的基因组在进化过程中经历了特殊的重排列和改变。这种现象在其他物种基因组中并不多见。该观点的提出源自一种简单的海葵种类——星状海葵 *Nematostella* 的基因组分析。

作为一种远古动物，海葵出现在寒武纪化石前期，比节肢动物（如三叶虫）和环节动物出现得还早。尽管海葵结构简单且年代久远，但是该物种仍然携带有一些在果蝇和线虫中早已经丢失的基因。更加不可思议的是，在星状海葵基因组中，大约有一半的遗传连锁是与人类基因组上极为相似，虽然这些基因在排列顺序上不尽相同（图 20-11）。拿人类第 10 染色体的 q24 区来说。在 4Mb 区域中一共包含有 11 个基因，其中有 *actin* 基因和 *SLK* 基因。*SLK* 是一个细胞分裂所需的激酶编码基因。在较小的星状海葵基因组中，它不仅包含有这 11 个基因，而且这些基因在 1Mb 区域中也是连锁的。这种局部同线性的保守性揭示出一种可能性，即连锁可能影响基因在某些细节层面的功能，当然目前还无法解释这种现象。通过更多的动物基因组测序，尤其是那些具有远古生物特性的物种，像海绵和扁形虫，将有望重建古动物的染色体组型，也即现在所有我们看到的当代多细胞动物种类祖先的确切染色体组型和遗传连锁结构。

用深度测序（deep sequencing）探索人类的起源

随着快速、廉价的大规模 DNA 测序技术的发展，那些在一年前几乎难以想象的实验今天已经成为可能。最近的一个例子便是穴居人基因组分析。

现代人类大约出现在 10 万年之前，在 50 万年前与穴居人具有共同的祖先。有证据表明，穴居人在 30 万年前消失之前，曾同现代人类共同生活在一定的区域范围内。有一种观点认为，这两个群体曾发生了交配，才导致现代人类基因组中携带有一些“穴居人基因”。为了证明这种可能性，科学家拟对穴居人的基因组进行完整的测序。

从保存完整的穴居人化石中，人们已经提取出其 DNA 样本。但是，这些样本受到了细菌和真菌 DNA 的严重污染。尽管如此，通过成千上万的短小 DNA 序列“读本”（参考第 21 章），从污染的 DNA 混合样本中分离了真正的穴居人基因组。实际上，在保存完整的穴居人化石中分离的 DNA，仅有 2%～3%属于真正的穴居人的 DNA，它们可以和猩猩、人类的参照序列相吻合。共有 1Mb 的穴居人基因组 DNA 序列被确认，这些序列 99.5%与现代人类基因组序列相同。穴居人同猩猩、人类的基因组比较说明，穴居人同人类间没有并合（comingling）的发生。现在，科学家正在对完整的穴居人基因组进行测序。对一个已经灭绝的动物基因组进行重新“复活”，这种想法实在是很了不起。

我们已经看到了基因组研究所揭示的多种基因组中编码序列和非编码调控序列的各种特性。现在，我们把目光转移到另外一个新生的学科中——系统生物学。藉此，我们希望能够更好地了解那些控制复杂调控过程的生物环路的特性，如动物发育过程。

系统生物学

分子生物学的成功，主要归功于它能够抓住那些简单的系统，以便能够研究其内在的详细机制。但是，同一些目标远大、有全局视野的研究策略相比，这些传统的研究方法开始褪去光环，这些新的策略采用定量及高通量的测量方法，通过模型构建、结构重组和理论方法的有机结合，替代传统的方法来研究更高级、更复杂的生物过程。这种交

叉学科研究的特性，恰恰定义了一个新的学科——系统生物学。系统生物学综合了数学、工程学、物理、计算机科学以及分子与细胞生物学等学科，以定量、可预测的方式来描述那些控制生命如何工作的相互作用环路的自然特性。系统生物学可以面对生物系统的许多不同层面，如信号传输、信号转导、细胞分裂和细胞骨架动力学等。作为基因的分子生物学教材，我们主要集中在从基因调控回路角度来阐明系统生物学。我们最大的期望是，藉此能够阐明那些靠传统的单成分分离研究所不能理解的基因调控原理。

系统生物学与另外一个新的学科——合成生物学紧密相关。就像系统生物学一样，合成生物学试图证明生物各环路的设计原理。然而，合成生物学是通过创造模拟的人工网络，模拟基因调控回路的一些自然特性，从而来证明原理。此方法使得我们能够检测调控系统是如何工作的。与自然界中存在的相对较为复杂的调控回路相比，这些人工网络相对简单，可以更精确地进行定量研究。

系统生物学所关注的不仅是基因调控的逻辑关系，而且还包括对进化内涵的解释。在第 19 章我们已经谈到，高等动物进化的主要驱动力来自于调控基因表达网络，而不是基因本身。例如，所有的动物都具有相类似的基因群，但它们表达的时空却不尽相同（有时候会截然相反）。换句话说，在进化过程中，网络调控是处于相对动态的，而受该网络调控的基因是处于相对静态的。

本节中，我们着重于从自然和人工构建的基因调控回路角度简要地介绍系统生物学与合成生物学。我们集中讨论遗传回路（genetic circuit）的设计原理和不同基因调控回路，而不过多地涉及该领域的数学细节。系统生物学是一门新生的学科，但是我们也要看到，该学科中的一些原理是从经典基因调控研究中衍生而来的，特别是我们在第 16 章所讨论的那些。但是，本章中我们将引入一些正式的语言，这些语言能把简单的例子拓展成一系列不同的生物调控系统。

含有节（node）与端（edge）的转录回路

调控回路可以通过一个含有节与端的简单网络来描述。节代表基因并用点来表示；端表示一个基因受另外一个基因产物调控，并用线表示（图 20-12a）。端可以通过引入方向来标明是否 A 基因调控 B 基因，或者相反。端还可以加标志来表示该调控是正调控还是负调控。因此，一条从基因 A 指向基因 B，末端为“⊥”结尾的线表示基因 A 的产物是基因 B 的负调控子（图 20-12b）。反之亦然，一条从 A 指向 B 的箭端线表示的是基因 A 的产物正调控基因 B 的表达（图 20-12c）。

我们现在从一个最简单的双节点例子开始，即基因 A 的产物调控基因 B 的表达（图 20-12a）。通过应答某个信号，基因 A 编码的调控蛋白质触发基因 B 的表达。该调控蛋白质也可能是一个抑制子。在这种情况下，转录过程通过一个诱发子触发启动，该诱发子能够灭活抑制子（图 20-12b）。或者，该调控蛋白质还可以是一个激活子，它能够应答信号分子，触发基因 B 的转录（图 20-12c）。

乳糖操纵子是受两个调控蛋白质调节的（图 16-6 中已经描述），是一个两种调控类型相互交织的例子：诱发子的存在触发转录起始，这一过程是通过灭活 Lac 抑制子，提高 cAMP 浓度，促进 CAP（分解代谢物活化蛋白质）激活子结合到 DNA 上面而触

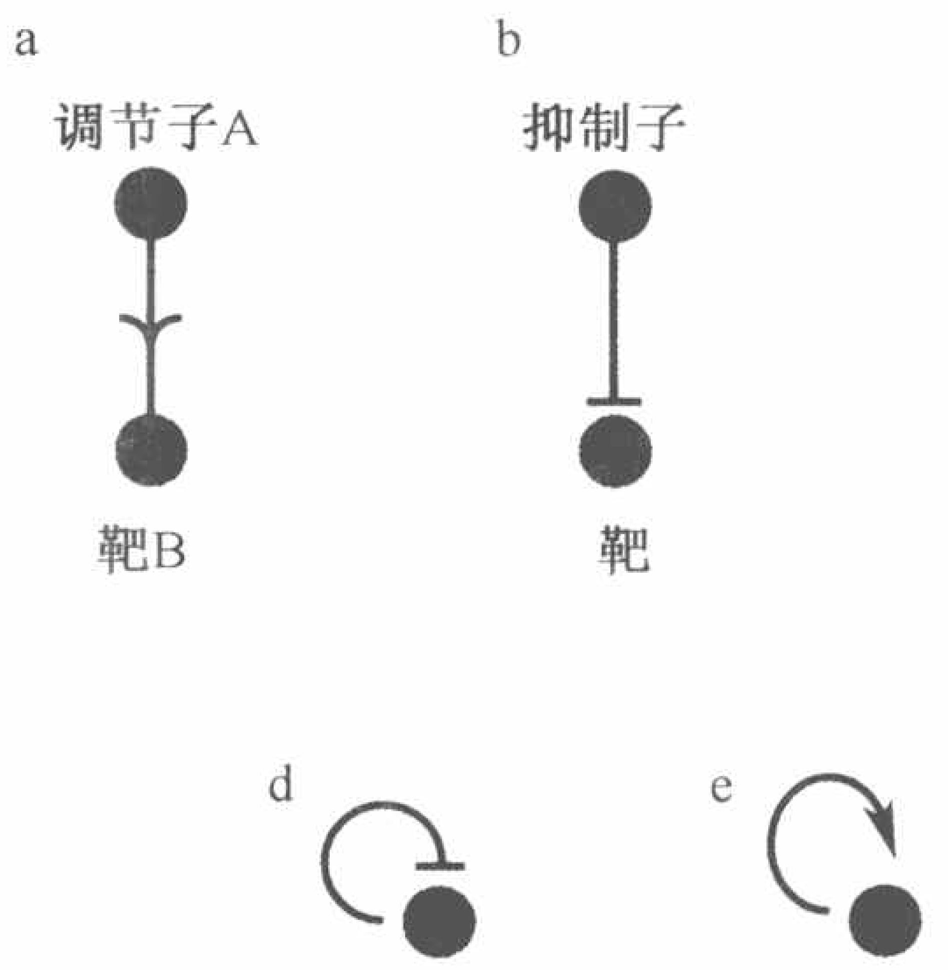


图 20-12　包含节与端的简单网络回路。(a) 一个简单的开关；(b) 负调控标志；(c) 正调控标志；(d) 自身负调控；(e) 自身正调控。

发的。因此，乳糖操纵子不是一个简单的开关：它的表达同时需要抑制子的消失和 CAP 结合到其 cAMP 配体上面。电子工程有一专业用语，“与门”逻辑关系，表示两个输入条件同时出现才能产生输出。在此例子中，这两个条件是抑制的取消和激活。与门可用图 20-13 的标志来描述。

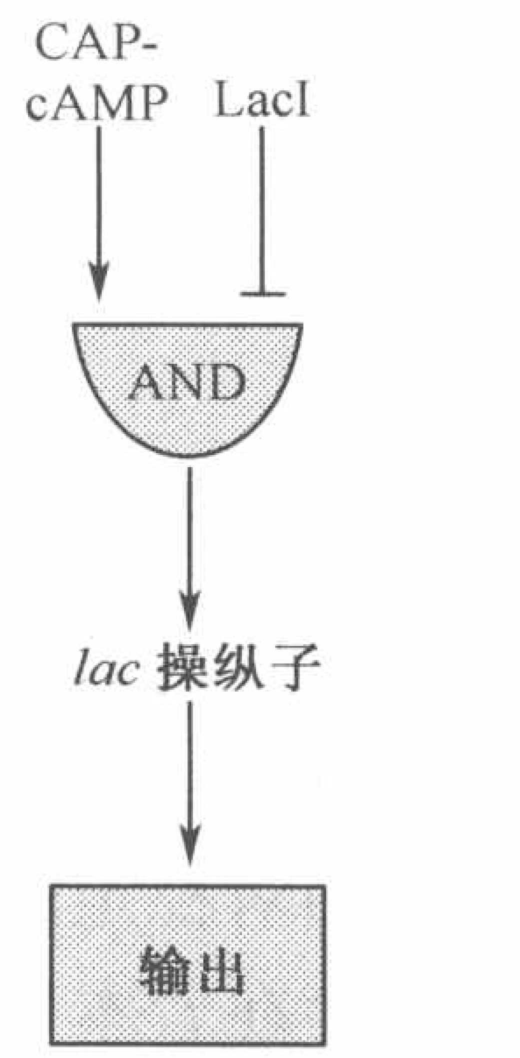


图 20-13　与门。乳糖操纵子受与门的调控。在该调控回路中，信号的输出需要 CAP-cAMP 的出现和 LacI 的取消。

自身负调控：降噪和降时

通常，调控基因除了调控其他靶基因的转录外，还能调控自身的转录，这种调控方式被称为自身调控，同样有正和负两种情况，每种情况都有自身的特性（图 20-12d，e）。我们先来考虑自身负调控，此时是一个编码抑制子的基因受其基因产物负调控。一个经典的自身负调控例子就是第 16 章中我们讨论的 λ 噬菌体的 cI 基因（需提示的是，cI 同样也是一个自身正调控的例子，对此我们会在下文讨论）。因此，CI 抑制子结合到操作子 O_{R3} 位点从而抑制自身的表达（图 16-27）。

自身负调控的生物学意义是什么呢？为什么在进化过程中它反复地被选择呢？其中一种解释已经在第 16 章中做了说明：自身负调控是保证调控蛋白质处于稳定状态的动态平衡机制。因此，CI 的水平下调到足够低来解除 cI 和其他靶基因的抑制，然后该基因转录的恢复将会提高细胞内抑制子的水平，最终恢复该抑制过程。相反，一旦基因的表达过大且抑制子的产物超过了所需要的量，那么自身负调控便会保证该基因处于沉默状态，然后抑制子的水平通过细胞生长与分裂，或者通过蛋白质酶降解，抑或是两者而最终下调。

自身负调控机制还有另外一个不太显而易见的优点：快速的反应时间。从尽可能快速且不会产生多余的抑制子的两个相对的角度考虑，如果使用一个强启动子，将能够保证产物的快速产生，但也会导致对稳定状态来讲过量的积累；从另一个角度考虑，如果使用一个相对较弱的启动子，也可以获得合适水平的抑制，但这需要较长的一段时间。

自身负调控可以同时满足两种需求：一个相对较强的启动子可以保证调控蛋白质的快速积累，而转录的自我抑制则能够保证在该抑制子达到一定的合适水平后自我关闭。数学模型和实验结果都确认了自身负调控比简单调控在实现同样的一种蛋白质积累水平上具有更快速的反应（图 20-14）。

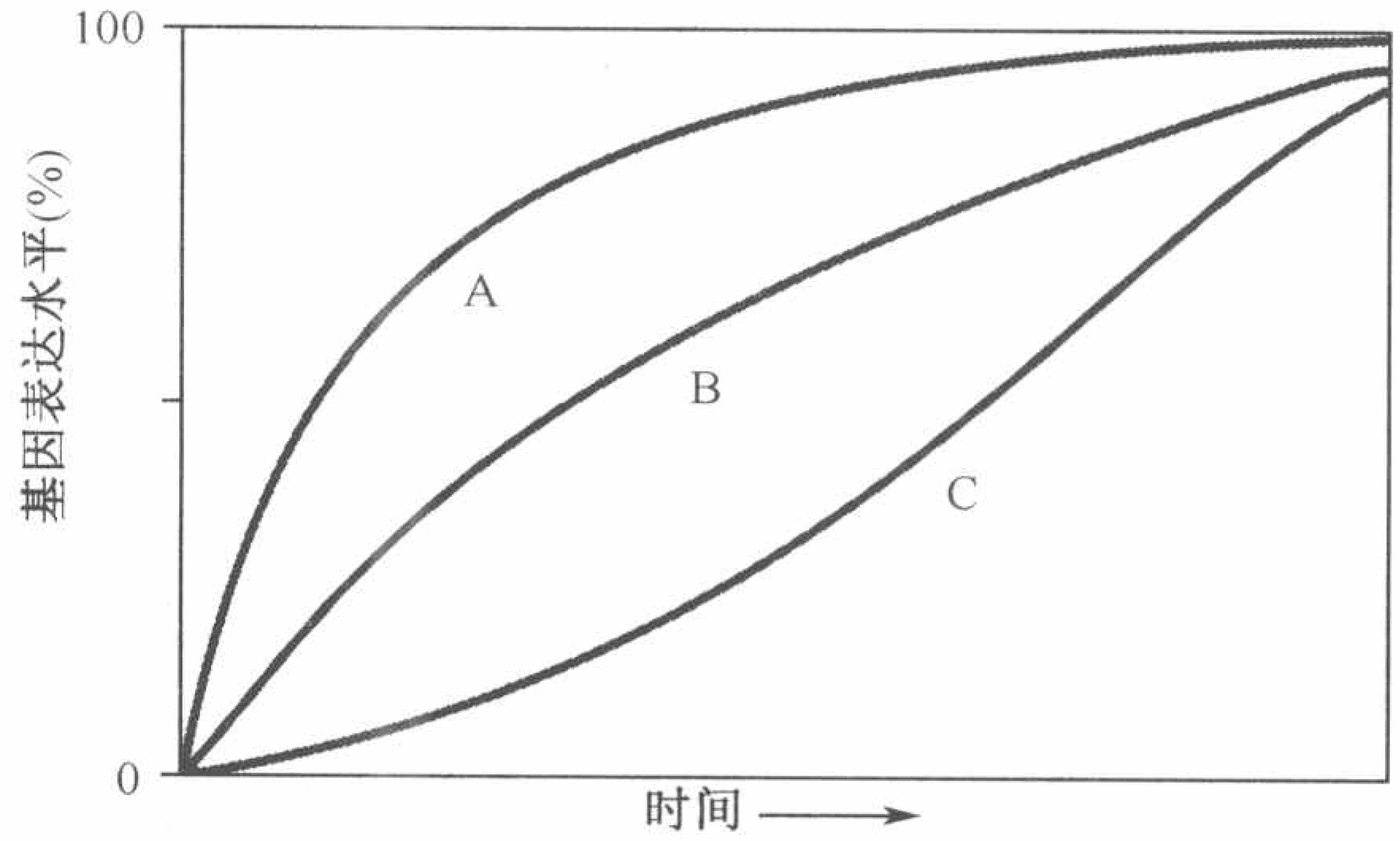


图 20-14 诱发子的动力学。(B) 简单开关、(A) 自身负调控开关和 (C) 自身正调控开关应答于一个诱导信号的不同动力学反应。

喧噪的基因表达

在讨论自身负调控过程中，基因表达噪声（noise）这一概念极为重要。直到最近，人们还是认为一个基因在同源的细胞群中的表达水平，在一个细胞中与另一个细胞中相比是相对稳定的。然而，我们现在意识到，基因表达水平在同一个细胞群的不同单个细胞中存在着本质上的差异，甚至在同一个细胞中，同一基因的两个拷贝本的表达也不相同。因此，我们在这里把“噪声”（noise）一词定义为是在基本相同的条件下基因表达的差异。噪声的存在表示随机因素影响着单个基因的表达水平。随机性（stochasticity）暗示着一个过程可以通过随机的尺度来划分为几个方面。正如我们所要看到的，一些调控基序是用来排除噪声的，而另外一些基序是用来放大它们的。

基因表达过程中的噪声来自于两大方面：内源性和外源性。这两种因素共同导致基因在一个群体中表达的不同。内源性噪声所指的是一个细胞内单个基因表达水平的不同是由于基因表达体中随机事件所引起的。一个经典的例子就是证明拥有两个同样拷贝基因的大肠杆菌内源性噪声的过程。该基因中一个拷贝连接一个能够编码红色荧光蛋白的报道基因，另一个拷贝与编码绿色荧光蛋白的报道基因连接。在没有内源性噪声作用的情况下，红色和绿色荧光蛋白的表达量应该是相同的，因此细菌应该显示黄色。相反，所观察到的结果是，许多细胞呈明显红色，还有一些明显为绿色（图 20-15a）。因此，每个基因在不同细胞内部表达并非一致，而是在一些细胞中一个拷贝的表达（如带有绿色荧光蛋白的拷贝）较另外一个拷贝高（如带有红色荧光蛋白的拷贝），而在另外一些细胞中，则恰恰相反。

外源噪声指的是，在一个几乎同源的细胞群中基因表达的差异，或者是在同一个细胞不同时间基因表达的改变。该噪声可能是由于每个细胞环境的微小的不均一性，或

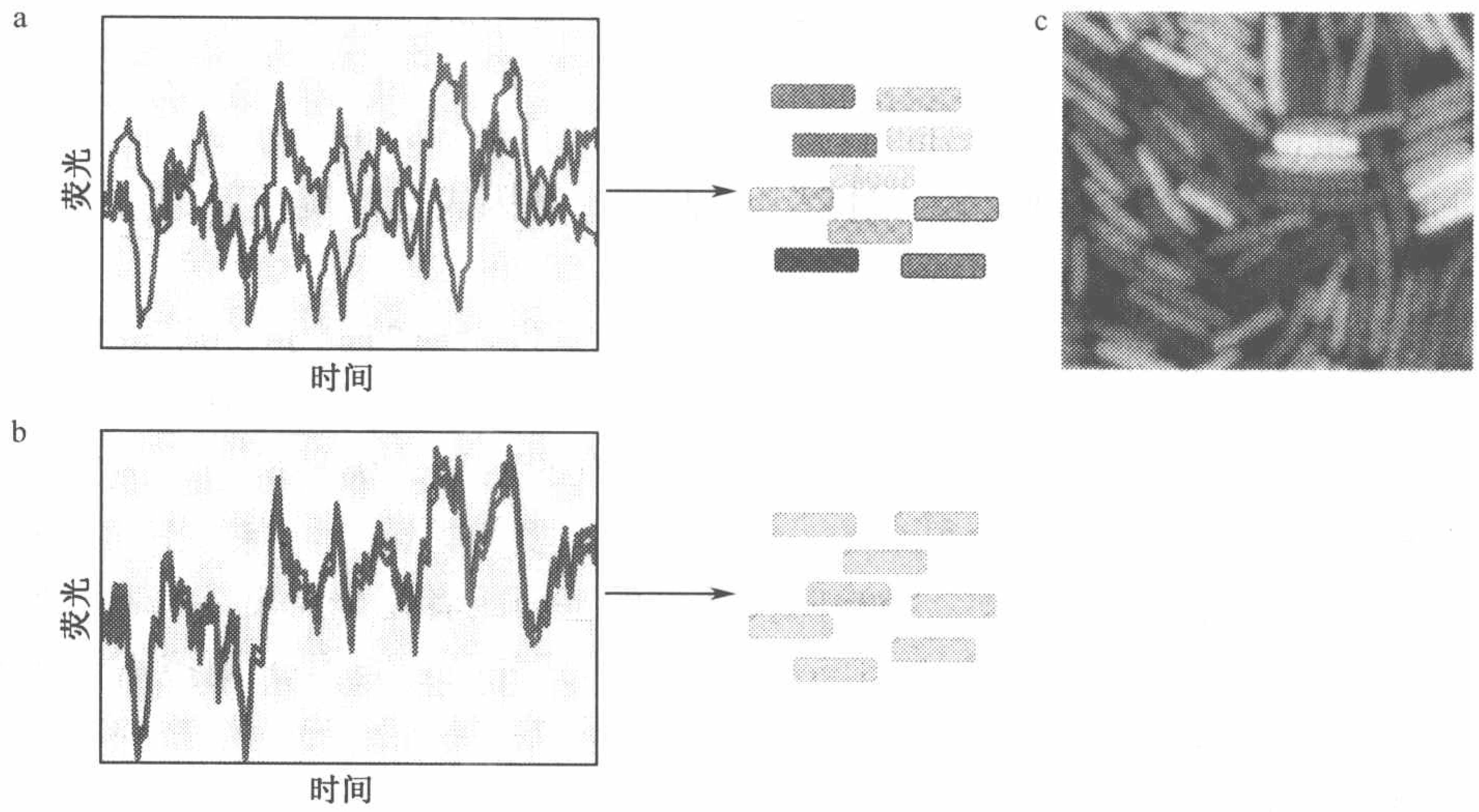


图 20-15 内源与外源噪声。在大肠杆菌细胞中，一个基因有两个拷贝，其中一个拷贝融合了一个红色荧光蛋白报道基因，而另一拷贝融合了一个绿色荧光蛋白报道基因。(a) 假设这两个基因在细胞内表达随着时间和个体各异（内源噪声）的预测结果。(b) 假设这两个基因的表达由于外源信号的作用而同步化的预测结果。(c) 荧光图片揭示内源噪声，除了黄色细胞之外，还有红色和绿色的细胞。(经允许，引自 Elowitz M. B. *et al*. 2002. *Science* 297：1184-1186.)

者是细胞在执行转录或者蛋白质合成的过程中容量的波动性所引起。一个外源性噪声的例子如图 20-15b 所示，两个基因的表达随时间变化。在这种情况下，两个基因的表达水平在单个细胞中呈现同步性：上升一段时期后便下调。这意味着单个细胞在承载一些或者全部基因的表达综合能力上是随时间发生波动的。从图 20-15c 显微荧光结果可以观察到，由于内源噪声的作用结果，其中一些细胞是黄色的，而另外一些细胞呈红色或者绿色。

重新回到一开始讨论的自身负调控，我们可以看到，为了应对噪声，该调控基序允许细胞补偿自调控基因的表达水平的变化。因此，调控 λ 噬菌体 CI 合成的自身负调控回路可以被认为是非常稳定的。稳定性（robustness）所指的是一个调控回路的输出对于一些特定参数是非敏感性。因此，CI 自控回路在克服 cI 基因表达过程中的噪声，获取稳定抑制子水平的能力方面是很稳定的。同时我们也应看到，其他一些调控基序也具有帮助细胞应对不同来源的随机因素影响的能力，例如触发基因表达的信号的波动。

随机性并非大肠杆菌所特有。实际上，该特征广泛存在于有生命的个体中。例如克隆猫的毛色，通过将体细胞核移植到已经去除核物质的卵母细胞而得到的克隆猫，与正常的核供体猫相比并不相同。由于克隆猫和核供体猫在遗传背景上是相同的，所以人们期望这些猫应该具有相同的毛色。但这并不意味着控制毛色的一系列遗传事件的前后关联不是完全的，而是存在一些随机过程。如同另外一个例子——同卵双胞胎的指纹不同。

自身正调控延迟基因表达

自身正调控过程发生在一个激活子蛋白质激活自身编码基因转录的时侯（图 20-12e）。同样的，λ 噬菌体 cI 基因的表达就是一个典型而复杂的例子：在细胞内 CI 浓度较低时，

CI 抑制子偏爱于占据驱动 cI 转录启动子（P_{RM}）上游的两个操纵子 O_{R1} 和 O_{R2} 的位置（图 16-26）。位于 O_{R2} 位置的 CI 蛋白质与 RNA 聚合酶相互作用而激活转录过程，以合成更多的 CI 蛋白质。当然，我们也可以看到，当 CI 达到一定的高水平后，该蛋白质还可以通过占据 O_{R3} 位置来抑制其转录过程。因此，cI 基因表达的调控是通过正和负自身调控机制来实现的。

现在让我们来介绍一个仅靠自身正调控的激活子蛋白质的例子（见图 20-12e）。基因产物的稳定性积累发生在如下两种平衡状态下：蛋白质合成与降解速率达到平衡（该过程应该是不稳定的），或者与细胞生长或分裂产生的稀释达到平衡时。因此，稳定性状态是指基因产物随时间的变化很小的一种状态。这里的关键是，自身正调控的基因从开启到实现稳定状态所需的时间，比自身负调控的基因或者完全无反馈的基因所需要的时间要长（图 20-14）。更确切地说，与其他调控形式相比，自身正调控需要更长的时间才能达到基因产物最大积累量的 50%。这是因为随时间而增加的生产速度，依赖于该激活子本身的积累。

自身正调控在一些缓慢发生的生物过程中非常有用，如发育过程就得益于与形态发生相关蛋白质的缓慢积累。又如，在枯草杆菌古孢子（或者原生孢子）发育过程中，控制晚期孢子形成的主要调控蛋白质（可变 RNA 多聚酶 σ 因子：σ^G 和 σ^K）能够刺激它们本身的结构基因以及与形态形成相关蛋白质编码基因的转录。因此，σ 因子和受它所调控的基因产物可以缓慢积累，因为 σ^G 和 σ^K 的产生依赖于它们自身的合成。

自身正调控还有另外一个优点。它是极端类型的调控，即被称为双稳开关的基础，对此我们将在下文阐述。

一些可锁定在不同稳定状态的调控回路

到现在为止，我们所介绍的调控回路都是可逆的，当打开某基因或者某些基因的信号被移除，该调控回路会重新返回关闭状态。然而，在一些例子中，当基因表达开关被启动后，该基因会锁定在启动状态持续较长的一段时间。这就是著名的双稳开关。控制枯草杆菌是否变成遗传感受态的回路就是一个被深入研究的双稳开关的例子。感受态是一种特殊的状态，处于该状态的细菌停止了增殖，但具有从其生长的环境中吸收裸露 DNA 分子，并通过遗传重组将同源序列整合到基因组上的能力。感受态的主调控因子是 DNA 结合蛋白质 ComK，该调控因子能够调控包括其自身在内的 100 多个基因（图 20-16）。维持此开关处于稳定状态的原因是多个 ComK 分子相互协作结合到 *comK* 基因启动子上。如我们在框 16-3 中 λ 抑制子（它本身就代表着一个典型的双稳事例，这点我们将在下文讨论）例子中所观察的那样，该类协作非线性地影响开关决定的输出，即与激活子的浓度相关。换句话说，输出的多少对 ComK 水平的变化非常敏感（反之则是不敏感）。

细胞能否开启 *comK* 的表达，是由 ComK 蛋白质降解稳定性水平上的调控网络所控制的。但是，最终决定 ComK 激活的因素是随机的。也就是说，在 ComK 不降解的情况下，群体中仅有一些细胞是处于感受态的。这可以通过一种带有荧光报道基因（绿色荧光蛋白基因）的细胞清晰地观察到，荧光报道基因指示的是 ComK 决定的基因活

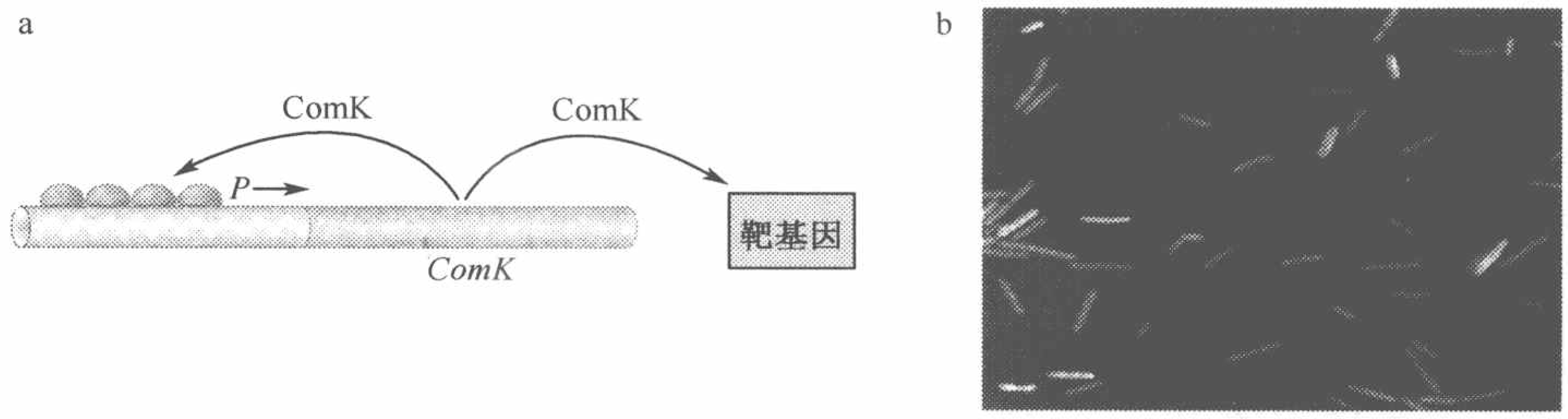


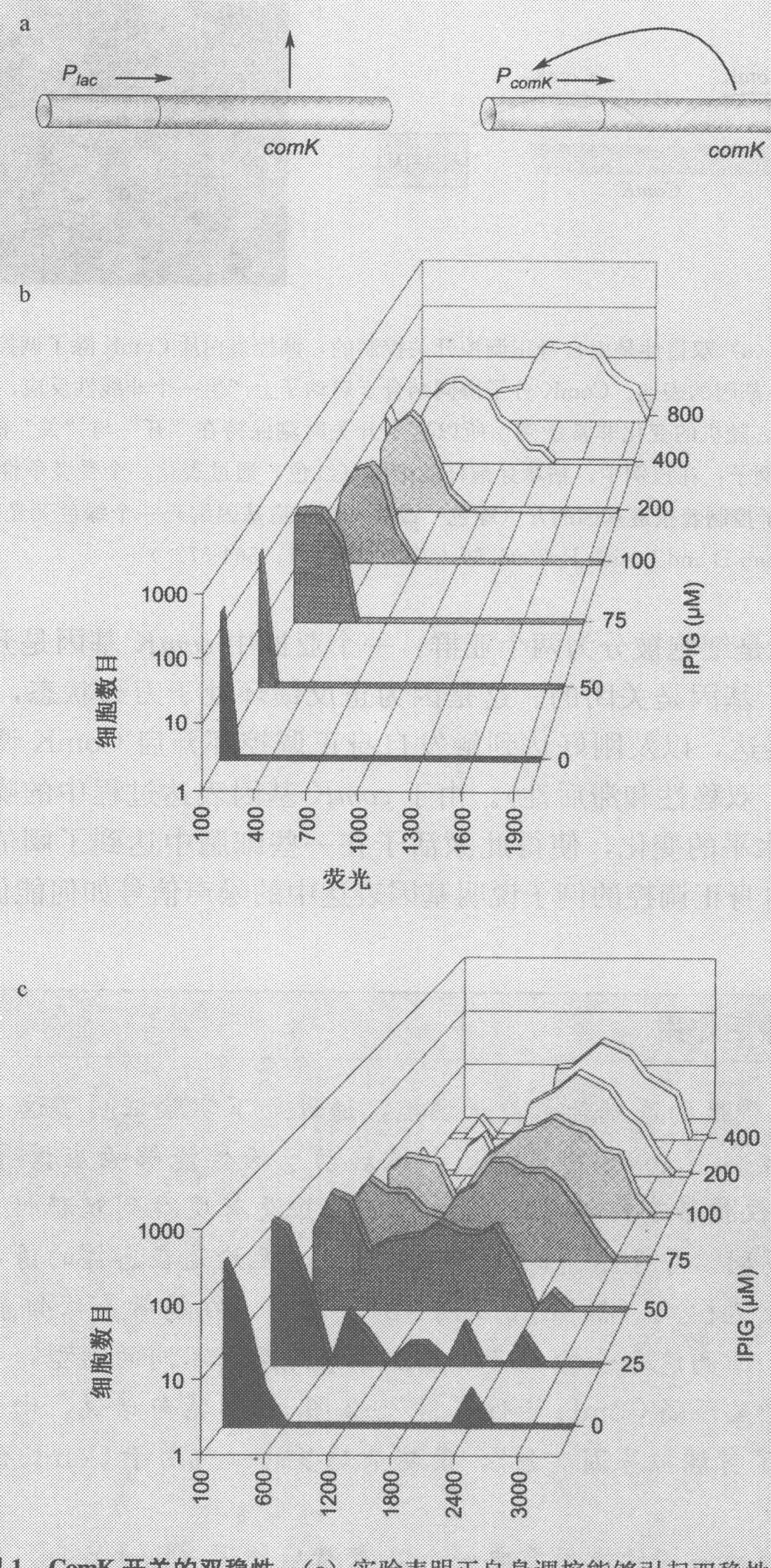
图 20-16 双稳性。(a) 双稳性是由自身正调控开关控制的，调控蛋白质 ComK 除了调控靶基因的表达外，还调控 *comk* 基因的表达。ComK 分子协同结合于启动子上产生一个非线性反应，使得其表达对 ComK 的微小的、随机的变化非常敏感。所以这个开关时刻保持在“开”与“关”的刀锋边缘。(b) 经典双稳性的例子：在该例中，枯草芽孢杆菌细胞（红色）通过表达一个受竞争性调控蛋白质 ComK 激活的启动子控制着报道基因的开（绿色）或关。此报道基因编码一个绿色荧光蛋白。(b 图经允许，引自 Dubnau D and Losick R. 2006. *Mol. Microbiol.* 61：564-572.)

性。图 20-16 显示的是细胞被分为两个亚群：一个亚群中 *comK* 基因是开启的，而在另外一个亚群中 *comK* 基因是关闭的。这是因为正反馈环处于刀刃状态，介于 ComK 的不足打开 comK 的表达，以及刚好达到触发自身正调控环开启 ComK 调控的基因二者之间（参考框 20-2，双稳性和滞后性）。由于 *comK* 基因表达过程中的噪声所导致的细胞间 ComK 蛋白质水平的变化，使得此激活子在一些细胞中达到了阈值浓度，而其他细胞中没有。这一自身正调控的例子说明基因表达中的噪声信号如何能促使细胞进入不同的状态。

框 20-2 双稳性与滞后作用

Comk 开关的双稳性的基础是自身正调控，通过如下实验证明了这一机制，用一个经修饰的在启动子控制下 *comk* 拷贝基因，该启动子活性能够响应诱导子上调或下调（框 20-2 图 1a）。在仅携带有修饰基因的细胞中，并没有观察到双稳性，Comk 决定的基因表达水平的提高同诱导子水平的提高基本对应，无论是在怎样的诱导子浓度中，其表达呈现出均一的表达模式（框图 20-2 图 1b）。然而，在携带有修饰基因和正常自身调控基因的细胞中，增高诱导子的水平导致细胞分为两种不同的类型，一种细胞呈现低 ComK 活性，另一种呈现高 ComK 活性（框 20-2 图 1c）。换句话说，由修饰基因所编码的 ComK 产物开启了自调控基因“泵”，使越来越多的细胞由于 ComK 的提高而处于开的状态。

严格来说，双稳性一词的使用需要一个具有滞后作用（hysteresis）特性的开关。滞后作用所指的是一种记忆状态，当一个开关由于某种条件打开后，即使撤去或反转该条件，开关也不会即时关闭。例如铁镁矿物材料，当这些物质在磁场中被磁化后，即使把磁场撤离，矿石还会保持磁性特征。现在我们回到原来的例子中——携带野生型和修饰型 ComK 基因的可响应诱导子的细胞。正如我们所看到的，添加越来越多的诱导子，将导致 ComK 水平不断提到，直至超出其阈值，导致自身正调控处于开的状态。我们来



框 20-2 图 1 ComK 开关的双稳性。(a) 实验表明正自身调控能够引起双稳性。示意图所示的是一个修饰的 *comK* 基因，该基因的启动子由一个受 IPTG 诱导的启动子所替代（就是 Lac 操纵子)。(b) 当细胞中只有修饰的 *comK* 基因时，comK 会在可被 IPTG 诱导的启动子控制下发生梯度应答反应 (c) 当细胞携带有完整的正自身调控和一个乳糖诱导 *comK* 拷贝基因时呈现双峰分布。注意：a 和 b 中细胞携带有一个 *comK* 启动子驱动的绿色荧光蛋白基因。(b，c 图经允许，引自 Maamar H. And Dubnau D. 2005. *Mol. Microbiol.* 56：615-624，Fig. 4E，J.)

看另外一种情况，即降低诱导子的水平，将使修饰型基因合成的 ComK 越来越少。我们还观察到，当诱导子的水平降低时，甚至当诱导子的水平低于开启 ComK 开关的浓度时，ComK 还是处于开的状态。也就是说，诱导开的因素都撤离后，ComK 仍然记得自己处于开的状态。

调控 λ 噬菌体溶原或裂解状态的开关也具有滞后作用。当原噬菌体受到 DNA 破坏试剂裂解细胞的诱导时，噬菌体能够反转进入裂解周期。也就是说，哪怕是当 DNA 破坏试剂撤离之后，该噬菌体也不再进入溶原状态（恢复合成 CI 抑制子）。作为一个相反的例子，由于乳糖的存在，乳糖操纵子开启，一但培养液中没有乳糖诱导子，乳糖操纵子则立即关闭。

自身正调控不是双稳性的唯一基础。一个具有两种变化状态的开关还可以通过相互抑制的途径来实现，即两种抑制子能够负调控对方的转录。如上所述，λ 噬菌体提供了一个经典的双稳性开关的例子，但该开关依赖的是一个双重负调控回路，而不是自身正调控；CI 与 Cro 抑制子的相互拮抗作用与病毒溶原与裂解状态紧密相关（第 16 章）。用系统生物学的语言来说，我们可以说 λ 噬菌体具有一个双向负端链接的双节点开关。

尽管许多双稳开关都是在细菌中发现的，但是这并不意味着双稳性仅局限于微生物。例如在胚胎发育过程中，线虫 *C. elegans* 产生两侧平行的味觉神经，分别称之为左侧 ASE 和右侧 ASE，分别表达不同的味觉受体基因。一个能够稳定处于一种或另一种状态的双向负反馈环，控制着一个普通的前体细胞究竟表达哪种受体。在这种状态下，开关的摆向并非随机的。确切地说，上游信号支配着开关的摆动方向，而双向负反馈环能够使其按预设的状态锁住开关。

前馈环——一种有优势特征的三节点网络结构

系统生物学领域中一个重要的贡献是，它从无数种理论上存在的、简单的调控回路中，发现只有一小部分能够在自然界中发生。显然，一些特定的环路拥有自然选择所偏爱的有利特征。一个典型的例子便是由三节点组成的环路（图 20-17a）。一共有 13 种可能的方法用箭头连接含有 3 个节点的网络环路。这些环路可以通过箭头的方向，箭头是否连接两个或三个节点，以及成对的节是否被一个或者两个箭头连接起来等相互区分。在这 13 种环路中，一个非常显著的环路是前馈环（图 20-17b），该环路普遍存在于自然界中。在这里，我们将其称为网络基序，这主要是因为该环路是遗传回路中的主要组成部分。前馈环基序包括转录因子基因 A 和受其调控的另一个转录因子 B（图 20-17b）。同时，这两个转录因子又同时调控该基序中第三个基因 C。注意：图 20-17b 中只是简单说明调控的方向（如节 A 调控节 B）而非符号。

如果符号（正或负调控）也加入方向端的话，那么就有八种不同的前馈环可以被区分开来。同样，自然选择还是偏爱其中两个，因此这两个比其他环路更常见。其中一个偏爱的前馈环基序（也被称为协调基序）是所有直接和间接的通路都是指向靶基因，代表着输出的方向，同时还具有相同的符号（例如 A 和 B 都是激活子）（图 20-17c）。

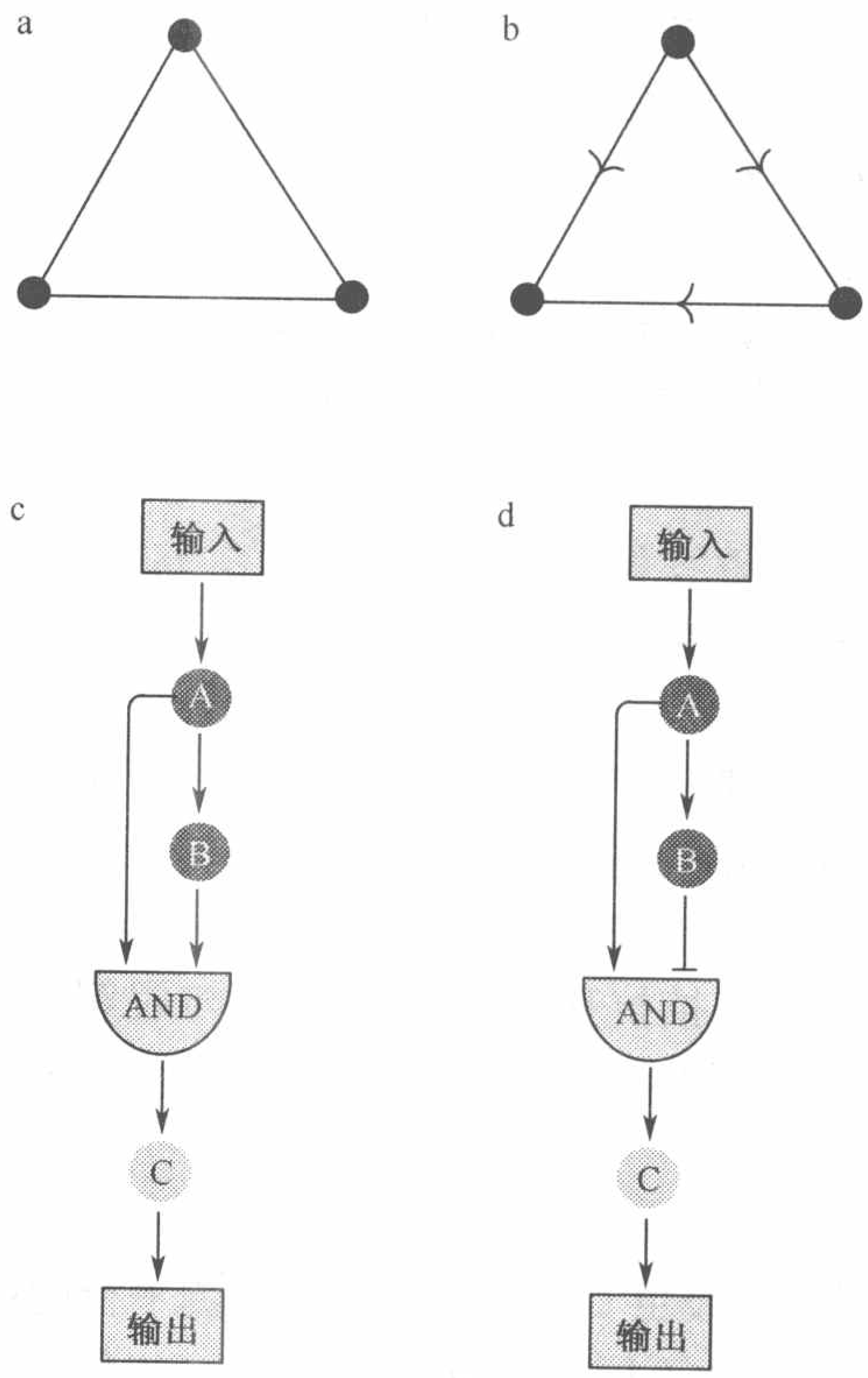


图 20-17 网络环路模型。(a)“三节点”网络环路，在该环路中每一个节点代表一个基因，这些基因通过三边而相连。(b) 前馈环，自然界中最为常见的网络环路。(c) 协调型前馈环，在该环路中，直接和间接的边都是靶基因的正调控信号。(d) 非协调型前馈环，在该环路中，直接作用边是正信号而间接作用边为负信号。

而另外一个偏爱的前馈环基序（也被称为非协调基序）是靶基因受到 A 通路直接正调控，B 通路间接负调控（图 20-17b)。在这两种基序中，靶基因的表达是受到逻辑与门的调控；也就是说 C 基因的转录在前一种环路中需要 A 与 B，而在后一种环路中需要 A 而非 B。

既然这两种基序在所有的前馈环，以及所有可能的三节点网络环路中最受偏爱，那么它们理所当然应该具有进化选择过程中最为基本的有利特性。的确如此，计算机模型与实验结果同时表明该基序具有使得它们在调控回路中更加有用的特征。例如，协调型前馈环具有要求持续的输入信号来维持靶基因 C 转录的特性（图 20-17c)。也就是说，这类型的前馈环是一个持续型检测子，它只对长时间维持的或者是稳定的信号发生应答。该特性源自以下的事实：靶基因的表达既依赖于第一激活子 A，又依赖于第二激活子 B 的有效积累。因此，输入信号必须持续足够长以满足第二激活子 B 达到激活靶基因 C 表达的阈值。换句话说，通过加入一个对输入信号反应的推迟，协调型前馈环帮助细胞分辨持续型真实信号以及随机的扰动（噪声)。

非协调性前馈环具有它本身的优势特性（图 20-17d）。它是一个脉冲发生器，导致基因表达在开启与关闭间切换。因此，激活子 A 打开靶基因 C 的表达，但当激活子 B 聚集到一定浓度后便将基因 C 的表达关闭。所以，当基因表达要求在一定的时间范围内时，此非协调型前馈环便发挥了作用。

发育过程中的前馈环

前面那些剖析阐明了基因调控复杂通路中的简单设计原理。在一些例子中，协调与非协调前馈环的组合导致产生各种各样的基因活动模式。其中一个引人注目的例子是孢子形成过程，该过程就是由上述一系列协调与非协调前馈环所形成的调控回路（图 20-18）。协调环保证环路的输入信号稳定，确保发育发生在正确的时间和地点。同样的，非协调环用来在形态建成过程中连续地产生基因表达脉冲。

此外，另外一个典型例子是果蝇胚胎背腹形成调控机制。我们在第 19 章中已经讨论过，母源调控蛋白质 Dorsal 启动这一过程，该蛋白质在胚胎中呈现宽范围的浓度分布。当 Dorsal 浓度处于中高或高时，它便直接作用于靶基因 *twist*，激活该基因的表达。Twist 也是一个调控蛋白质，它与 Dorsal 协调作用，调控一系列基因的表达，例如 *snail*。因此，该调控基序是一个明确的协调型前馈环例子。然而，除此之外，*snail* 编码一个转录抑制子，许多 Dorsal 和 Twist 调节的靶基因受它的抑制。因此，这类靶基因是受非协调型前馈环所调控的。所以说，在细菌孢子形成过程中，由 *dorsal*、*twist*、*snail* 以及下游基因所构成的环路是由协调和非协调型前馈环组成的。在果蝇胚胎发育过程中，背腹形态形成过程受前馈环的调控。因此，在中胚层，由于 Dorsal（及 Twist）及 Snail 的表达水平较高，所以由 Snail 介导的靶基因的抑制被关闭，而在外胚层神经元，由于 Dorsal 和 Snail 的水平较低，所以由 Snail 介导的靶基因的抑制是打开的。

有些调控回路产生节律性的基因表达模式

对基因表达的调控我们一般考虑打开、关闭或者是调控表达水平。但是，在生物体内还广泛存在着另外一种基因调控模式——节律性，即大部分基因的表达都是在一个规则的时间间隔内呈现周期性上调和下调。定量地阐明调控节律性行为的调控回路是对系统生物学的一个重要挑战。

一个相对简单的节律性调控回路的例子是新月柄杆菌（*Caulobacter Crescentus*）的细胞周期（图 20-19）。在此例中，中枢调控子（CtrA 和 GcrA）的水平以周期性模式交替升高和下降。它们的交替出现驱动其他基因在细胞周期过程中呈现节律性表达模式。

生物钟是一个众所周知的节律性行为的例子，在生物钟的调控下，大量基因在日夜交替周期的不同时间呈现节律性表达。在果蝇和哺乳动物中，其昼夜节律部分地受到负前馈环的调控，其中激活子蛋白是 Clock 和 Cycle，抑制子是 Per（Period）。Clock 和 Cycle 蛋白结合到 *per* 基因调控区域，激活该基因的转录。当 Per 蛋白在细胞中累积到一定的水平后，它能够反作用于 Clock 和 Cycle，从而关闭自身合成。一旦 *per* 基因的转录被关闭后，由于蛋白酶降解造成的不稳定性，Per 蛋白将慢慢从细胞中被清除。这

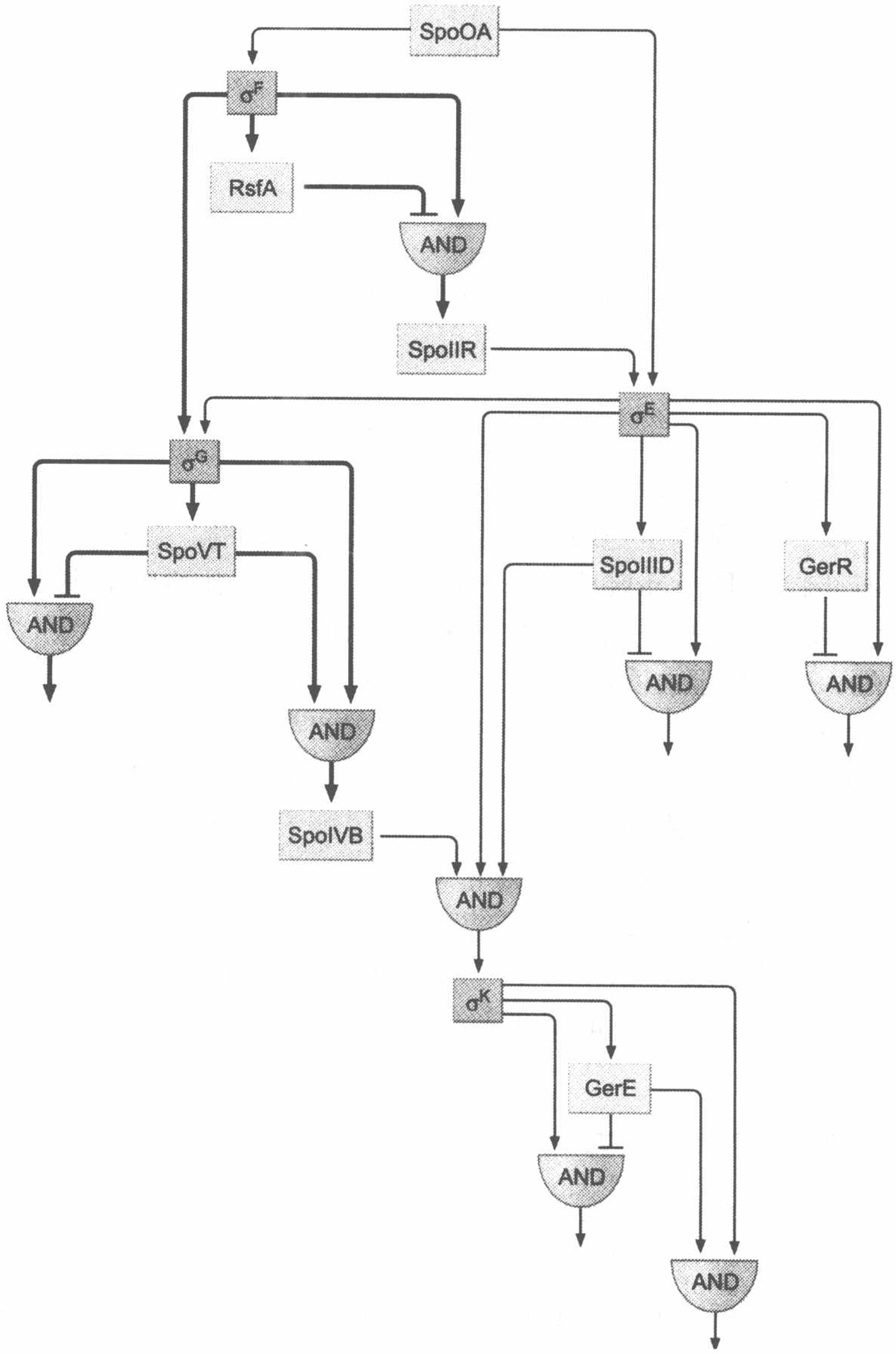


图 20-18 一系列前馈环组成的环路调控孢子形成过程。图中名字所表示的是调控或信号蛋白质。为了简洁，自身正调控回路因子 σ^G 和 σ^K 没有在图中画出。（经允许，由 Wang S. T. et al. 2006. *J. Mol. Biol.* 358：16-37，Fig. 5. 重新绘制）

将导致抑制子恢复到一个次级阈值的水平，即 Per 蛋白没有足够高的浓度来抑制 Clock 和 Cycle 的活性。由此，*per* 基因的表达重新打开。*per* 基因表达的开/关周期有助于确定基因活性的 24h 循环过程。Per 蛋白的合成与降解的时间是关键的依赖因素。Per 蛋白稳定性的改变能够改变节律的频率，甚至产生不正常的以 22 或 26h 为周期的循环，而不是正常的 24h。然而，关于生物钟为什么稳定地保持 24h 周期的机理还未被完全理

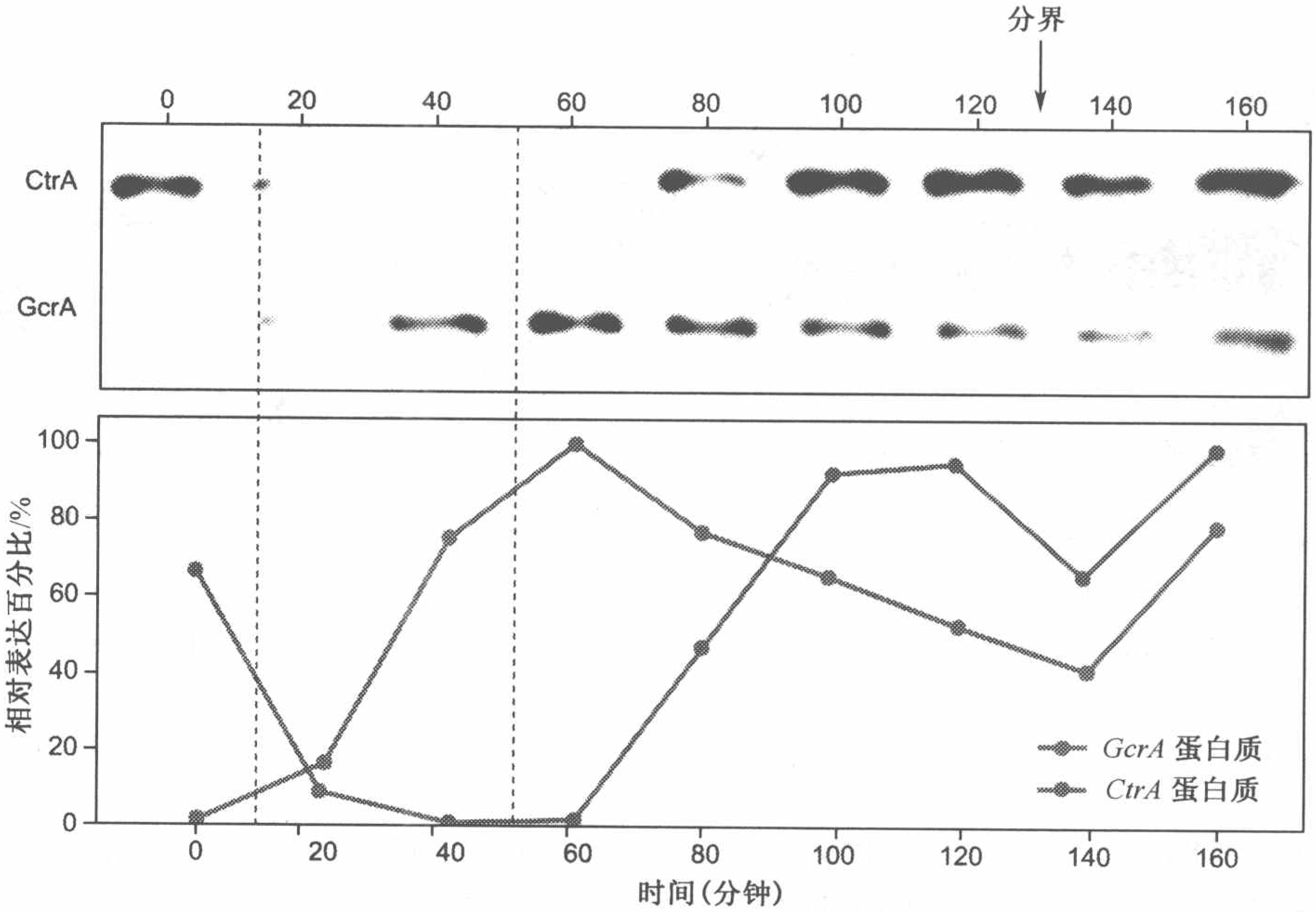


图 20-19 调控子 CtrA 和 GcrA 在柄杆菌属细胞周期的过程中的波动表达。（经允许，引自 Holtzendorff J. et al. 2004. *Science* 304：983-987，Fig. 3B，C.）

解，毫无疑问，仍有一些确切存在但尚待阐明的机制。

有趣的是，自身负调控还涉及另外一个与周期基因表达毫不相关的例子：脊椎动物胚胎中的体节形成过程。体节是胚层细胞的浓缩块，它们将形成重复肌肉节和脊椎动物的脊柱（图 20-20a）。体节从头到尾的形成模式依赖于调控基因 *her1* 和 *her7* 有节律的活性的开启与关闭，至少在斑马鱼中是如此。这些基因在此后的体细胞中保持周期性表达，直到这些细胞准备分化形成物理性体节为止。随着每一批新细胞的成熟，其节律性终止，从而导致一些细胞停留在周期的峰顶，另外一些停留在峰谷，并在空间顺序上体现了体节的形成。紧随新体节形成的是下一组将来形成体节的细胞重复相同的步骤，包括新一轮的基因表达的开启和关闭，并最终形成新体节。

her1 和 *her7* 在斑马鱼中节律性表达的开启和关闭已经应用于数学建模和计算机模拟。Her1 与 Her7 蛋白被认为结合到 *her1* 和 *her7* 基因的调控区域从而关闭转录过程。但是该抑制过程仅仅维持短暂时间：由于蛋白酶的降解，一旦 Her1 与 Her7 蛋白质降到一个关键的阈值水平后，转录过程重新被恢复并开始一轮新的循环：蛋白质合成开始—抑制状态恢复等。抑制基因产物的节律水平调控其他基因的表达，从而确定出新体节的形成模式。在斑马鱼中，每 30min 形成一个新的体节。解释体节形成的周期长度的一个关键点是 *her1* 和 *her7* 基因启动转录同自身抑制子积聚到有效浓度以抑制自身合成过程的时间间隔（图 20-20b）。

Her1 与 Her7 自身调控环完美地阐述了体节基因如何在单个的细胞中呈现开/关表达的循环模式，但是这些基因是如何在即将形成体节的一个细胞及其相邻的细胞中保持

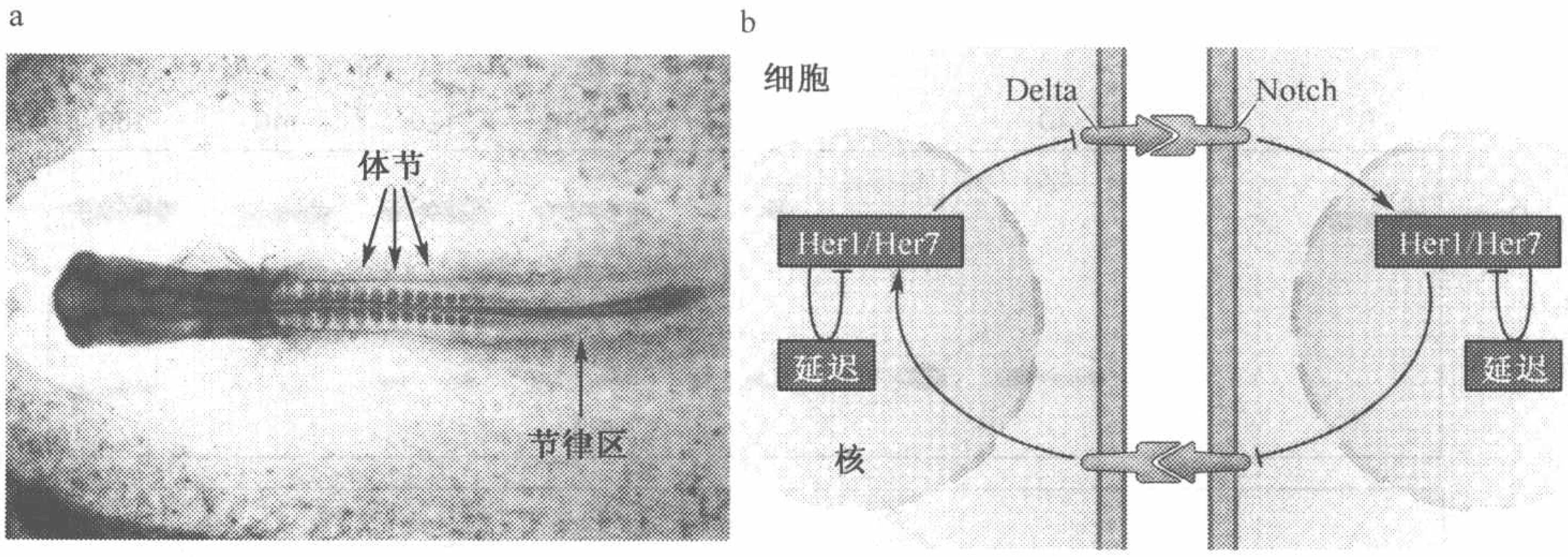


图 20-20 体节基因在脊椎动物发育过程中的表达。(a) 图中所示为鸡胚胎发育过程中的体节。箭头所示的是体节，节律区是节律基因表达的区域，新的体节将在此产生（图片由 Julian Lewis 提供）。(b) 图中所示为斑马鱼体节基因节律表达的产生和同步模型。节律基因的表达受负反馈环的调控，其中涉及自身抑制子 Her1 和 Her7。不同细胞间节律的同步化是通过 Delta/Notch 信号通路来完成的。Her1/Her7 抑制 Delta 配基的生成。相反的，Notch 信号刺激 Her 蛋白质的表达。（摘自 Julian Lewis 提供的图片。）

同步化的周期性表达呢？同步化是通过细胞与细胞间的信号通路来完成的。Her1 与 Her7 除了抑制其自身外，它们还抑制一个细胞膜编码蛋白质 Delta 基因的表达。Delta 结合到与其相邻细胞的细胞受体蛋白质 Notch，并激活 Notch 信号系统，进而激活编码 Her1 与 Her7 蛋白所需基因的表达。当开/关循环中 Her 蛋白水平较高时（即 her 基因表达较低），产生比较少的 Delta 配基（图 20-20b）。这将导致与其相邻的细胞中 Notch 信号系统的活性与 her 基因的表达都比较低。相反的，当 Her1 同 Her2 蛋白水平较低时（即 her 基因表达较高），细胞中 Delta 含量便升高，从而激活相邻细胞 her 基因的表达。对于每一种情况，信号从相邻细胞通过 Notch 与其内 Her1 和 Her7 蛋白相互协作，从而保持细胞与其相邻细胞节律之间的同步性。因此，细胞内自身调控与细胞间信号传递共同使得体节构成细胞呈现节律性行为。

用人工合成的环路模拟自然的调控网络

理解主导调控回路设计原理的一个辅助方法是：建立一些相对简单的环路，并使这些环路能够模拟自然系统的一些特性，这就是合成生物学领域的目标。环路设计的一个成功的典型例子是抑制振荡器，该实验加强了我们对节律性的讨论。抑制振荡器是一个在大肠杆菌中生成的三节点网络环路，该环路由三个相互连接的调控蛋白组成一个环形的模式，三个端的连接符号都为负号。此抑制子由三种细菌抑制蛋白 λCI、LacI 和 TetR 编码基因所组成，CI 蛋白抑制 LacI 编码基因，LacI 进而抑制 TetR 编码基因，而 TetR 反过来抑制 CI，从而形成整个网络环路。人们可能会推测，此类三节点环路可能会导致这三个基因的转录都处于一个低水平的稳定态。但恰恰相反，此抑制振荡器呈现出一个强烈的转录节律模式，其周期大约为 2h。理论上讲，基因表达过程中噪声导致这三种抑制蛋白水平产生波动，这将阻止系统达到稳定态，取而代之的是一个节律表达模式。但是此抑制振荡器与上述的自然系统相比，其节律行为的稳定性依然相差甚远，这说明了一个事实，人工合成的环路在模拟自然界中的复杂（尚未被充分阐明的）环路

时还具有许多不足之处。

通过合成还产生了许多其他网络环路，这些环路呈现出各种各样的行为模式。其中一个例子是人工回路文库，该文库由许多转录因子和启动子通过不同的组合所构成。该回路文库中的回路能对不同的输入信号组合产生不同的应答。另外一个例子是来自“发射子”与“应答子”菌株，这类菌株能够在琼脂板上产生基因表达的条纹模式。发射菌株生长在培养板的中央，它能够产生信号分子，并从培养板的中央向四周扩散形成一个浓度梯度。两种不同的应答菌株分散长在同一个培养板上，这两种应答菌株对发射菌产生的信号分子浓度的高低应答不同，可通过产生可辨别的发光报道蛋白来区分。因此，一种应答菌株在距离发射菌较近处产生一个有色光环，而另外一种应答菌株在距离发射菌较远处产生另外一个有色光环。

展望

系统生物学的方法使系统辨别某一复杂细胞过程中所涉及的每一个成分成为可能。获得此类信息能力的提高，推动了生物学家数据分析方法的彻底转换。我们现在可能的问题是“该生物为什么形成这样的模式”，而不再是“这个过程是如何工作的”。展望未来，随着系统生物学与越来越完善的合成生物学的有机结合，由此引申出来的理论将有可能在未来创造出一个能够自我繁殖的有最小回路的人造细胞。到了那一天，我们将量身定做人工细胞，使其拥有有效代谢污染物、重新利用废弃物品、将太阳能转化为燃料或者治疗人类疾病等能力。

小结

通过大规模 DNA 测序，许多动物的完整基因组序列已经被确定，包括果蝇、海鞘、鱼类、小鼠和人类等。借助一系列实验和计算机的方法，蛋白质编码基因、非编码基因（例如那些特异性 microRNA）和 DNA 调控序列可以被系统性地识别出来。全基因组覆瓦式芯片（whole-genome tiling array）能高效地识别转录序列和整个基因组的内含子与外显子。而 ChIp-chip 芯片对识别 DNA 调控序列更加有效，例如组织特异性表达的增强子序列。

目前为止，比较基因组分析给我们带来最惊讶的结果是：各种相差甚远的动物间基因组却包含一系列相似的基因。例如低等动物海葵与人类都有一个相似的遗传“工具箱”。工具箱中所包含的是一些指导不同类型细胞发育的基因，包括细胞信号通路中 FGF、Notch、EGF、Wnts、Hedgehog 和 TGF-β 等分子。调控基因也是一群保守基因，如 Hox 复合体所编码的基因以及关键的转录因子，这些转录因子含有全长的 DNA 结合域，如 HLH、bZIP、HGM 和锌指结构域等。最近，海葵与人类基因组比较分析发现这两个物种之间具有明显的同线性。所以不仅是两个基因组含有相似的基因，而且最少有一半的基因在保守的连锁群中。

更加高效的新测序技术的发展为新的重大发现敞开了大门。现在，重新构建灭绝物

种的基因组已经成为可能。从穴居人化石中提取的 DNA 可以进行测序获得基因组信息。在未来几年中，重新构建出一个完整的穴居人基因组是有可能的。

系统生物学是一门新生学科，该学科综合定量与高通量的测量方法、建模、重组和理论知识，最终阐明生物机体内不同层次的内在机制。当涉及调控回路时，系统生物学试图揭示那些把每个元件都提取出来的方法不可能知道的基因控制的原理。合成生物学也试图阐明设计的原理，为了实现这一目标，所采用的途径是创造人工调控网络模拟天然回路的特征。

转录环路是由节点和端边构成的，节点表示的是基因，端边表示一个基因被另一个基因所调控。一个简单的双节点调控基序所表示的含义是一个基因调控另一个基因的表达，调控可以是正的，也可以是负的。另外一个简单的基序是自身调控，也就是基因调控自身的表达。自身负调控指的是一个基因的产物抑制该基因本身的表达，它具有抵御噪声的特性，所谓噪声是指几乎均一的条件下基因表达的波动性。而自身正调控可使基因表达缓慢增加达到一个稳定水平的特性。自身正调控的一个极端的例子是双稳开关环路，其中一个基因的表达保持在开或关的位置相当长的一段时间。

在调控网络环路中，另外一个常见的基序是前馈环。前馈环是一个三节点基序，它所表示的是调控（基因 A）能够控制靶基因和第二个调控基因（基因 B）的表达。而第二个调控基因 B 也调控靶基因的表达。因此在前馈环基序中，基因 A 能够直接调控或者通过基因 B 间接调控靶基因的表达。靶基因的表达可以说是受到一种“与门”的调控，有两种情形：第一种是两个激活转录因子同时存在，另外一种情形是激活子的存在和抑制子的消失。

在自然界中，一些调控回路能够使一些基因的表达产生节律性循环，例如细胞周期、发育和昼夜节律。该环路的设计原理是：其中一个调控蛋白的出现能够导致其自身的消失和另外一个调控蛋白的出现。而第二个调控蛋白的出现又反过来抑制其自身的表达和促进第一个调控蛋白的表达，从而构成一个连续的基因表达开/关循环。人工合成的网络已经成功将三个抑制子通过随机的方法连接成一个圆形的调控回路，可以模拟出基因表达的循环模式，但还达不到自然存在的节律控制子的稳定性。

参考文献

书

Alon U. 2006. *An introduction to systems biology: Design principles of biological circuits.* Chapman & Hall/CRC, Boca Raton, Florida.

Sussman H.E. and Smith M.A., eds. 2006. *Genomes.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

基因组

Arnosti D.N. 2003. Analysis and function of transcriptional regulatory elements: Insights from *Drosophila. Annu. Rev. Entomol.* **48:** 579–602.

Berman B.P., Nibu Y., Pfeiffer B.D., Tomancak P., Celniker S.E., Levine M., Rubin G.M., and Eisen M.B. 2002. Exploiting transcription factor binding site clustering to identify *cis*-regulatory modules involved in pattern formation in the *Drosophila* genome. *Proc. Natl. Acad. Sci.* **99:** 757–762.

Frazer K.A., Pachter L., Poliakov A., Rubin E.M., and Dubchak I. 2004. VISTA: Computational tools for comparative genomics. *Nucleic Acids Res.* **32:** W273–W279.

Green R.E., Krause J., Ptak S.E., Briggs A.W., Ronan M.T., Simons J.F., Du L., Egholm M., Rothberg J.M., Paunovic M., and Pääbo S. Analysis of one million base pairs of Neanderthal DNA. *Nature* **444:** 330–336.

Osoegawa K., Mammoser A.G., Wu C., Frengen E., Zeng C., Catanese J.J., and de Jong P.J. 2001. A bacterial artificial chromosome library for sequencing the complete human genome. *Genome Res.* **11:** 483–496.

Putnam N.H., Srivastava M., Hellsten U., Dirks B., Chapman J., Salamov A., Terry A., Shapiro H., Lindquist E., Kapitonov V.V., Jurka J., Genikhovich G., Grigoriev I.V., Lucas S.M., Steele R.E., Finnerty J.R., Technau U., Martindale M.Q., and Rokhsar D.S. 2007. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* **317:** 86–94.

系统生物学

Alon U. 2007. Network motifs: Theory and experimental approaches. *Nat. Rev. Genet.* **8:** 450–461.

Bintu L., Buchler N.E., Garcia H.G., Gerland U., Hwa T., Kondev J., and Phillips R. 2005. Transcriptional regulation by the numbers: Models. *Curr. Opin. Genet. Dev.* **15:** 116–124.

Bintu L., Buchler N.E., Garcia H.G., Gerland U., Hwa T., Kondev J., Kuhlman T., and Phillips R. 2005. Transcriptional regulation by the numbers: Applications. *Curr. Opin. Genet. Dev.* **15:** 125–135.

Crosson S., McAdams H., and Shapiro L. 2004. A genetic oscillator and the regulation of cell cycle progression in *Caulobacter crescentus*. *Cell Cycle* **3:** 1252–1254.

Dubnau D. and Losick R. 2006. Bistability in bacteria. *Mol. Microbiol.* **61:** 564–572.

Endy D. 2005. Foundations for engineering biology. *Nature* **438:** 449–453.

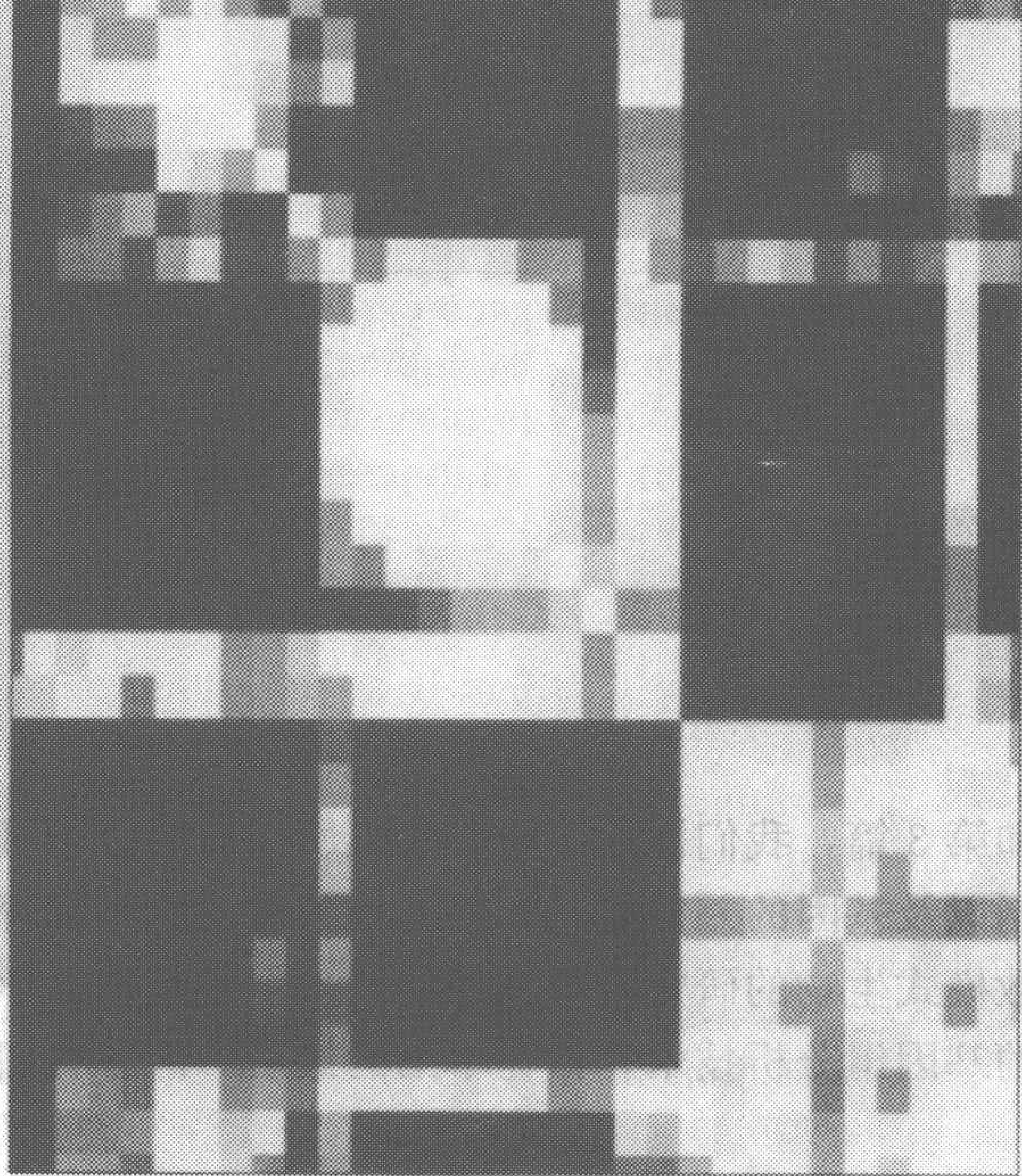
McAdams H.H., Srinivasan B., and Arkin A.P. 2004. The evolution of genetic regulatory systems in bacteria. *Nat. Rev. Genet.* **5:** 169–178.

McGrath P.T., Viollier P., and McAdams H.H. 2004. Setting the pace: Mechanisms tying *Caulobacter* cell-cycle progression to macroscopic cellular events. *Curr. Opin. Microbiol.* **7:** 192–197.

Raser J.M. and O'Shea E.K. 2005. Noise in gene expression: Origins, consequences, and control. *Science* **309:** 2010–2013.

Sprinzak D. and Elowitz M.B. 2005. Reconstruction of genetic circuits. *Nature* **438:** 443–448.

Vilar J.M., Guet C.C., and Leibler S. 2003. Modeling network dynamics: The *lac* operon, a case study. *J. Cell Biol.* **161:** 471-476.



第5篇 方法

本篇概要

- 第21章 分子生物学技术
- 第22章 模式生物

在本书的第2篇和第3篇，我们概述了中心法则所代表的分子机制，第4篇则集中讲述了基因表达的机制以及基因的差异表达如何调控各种动物的发育和进化的机制。这些知识大多源于对少数模式生物的研究，所使用的技术主要是遗传学、分子生物学和生物化学技术以及最近的基因组分析技术。作为本书的最后一部分我们将扼要地介绍研究方法和模式生物。

第21章简要讲述了分子生物学和生化基本技术。这些技术可以从细胞中分离（分离就是将这些分子从混合物中分开）各种分子（DNA、RNA和蛋白质），从而可以在体外研究其纯合物。第22章概述了少数模式生物的基本特征，它们的研究促进了现代生物学的发展。这些模式生物包括噬菌体和细菌、酵母、秀丽新小杆线虫、黑腹果蝇和小鼠。人们通过这些生物的遗传分析可以研究体内生物学过程。体内遗传学方法和体外生化技术的结合使分子生物学得到了飞速发展，并在过去的50年里使我们对生物学的理解有了革命性的变化。

当人们可以分离代表某一基因的特定DNA片段后，分子生物学进入了一个黄金时代。早期，人们就能从生物体中获得大量DNA，但直到20世纪70年代中期才发明了分离特定基因的方法。第21章我们将首先讲述如何应用限制性内切核酸酶和电泳来分离特定DNA片段，接下来是如何在体内扩增和表达这些片段。

然后，我们将介绍用聚合酶链反应进行DNA体外扩增的相关技术以及DNA测序技术，这些技术都需要使用化学合成的DNA片段作引物。因此我们也会简要说明DNA片段的化学合成技术。即使从一个DNA分子开始，PCR也能纯化并无限复制任一DNA片段。PCR扩增技术使很多学科领域发生了根本性的变革，包括法医学、医学、生态学，当然还有分子生物学。

在20世纪70年代中期和80年代，DNA测序还需手工进行，并相当繁琐。90年代在人类基因组计划这一宏伟目标的激励下，人们研制出了高度机械化的DNA测序方法，现已发展到数天或数周内就能完成一个基因组的核苷酸测序。

第21章的第一部分描述了运用双脱氧链终止法的测序策略及其近期在高通量测序上的运用。这一部分以对全基因组测序的令人激动的新方法和超高通量的纳米技术测序法的介绍而结束。在这一章的第二部分，我们将处理蛋白质纯化和分析的方法并简要描述蛋白质组研究这一新兴领域。本章的第三部分描述了研究蛋白质和核酸相互作用的研究方法，如DNA足迹法、凝胶迁移率变动分析技术法和染色质免疫共沉淀方法。

第22章我们讲述了若干模式生物，主要着眼于研究人员使用这些简单生物对感兴趣的问题进行研究的基本原则。按照基因组复杂性和生活周期长短来分，最简单的生物

要数细菌病毒，即噬菌体了。细菌和噬菌体的研究揭示了很多DNA功能的基本特征，包括基因表达的诱导、DNA复制、重组和修复。大肠杆菌（*E. coli*）及其噬菌体是20世纪60年代早期阐明遗传密码的关键生物材料。

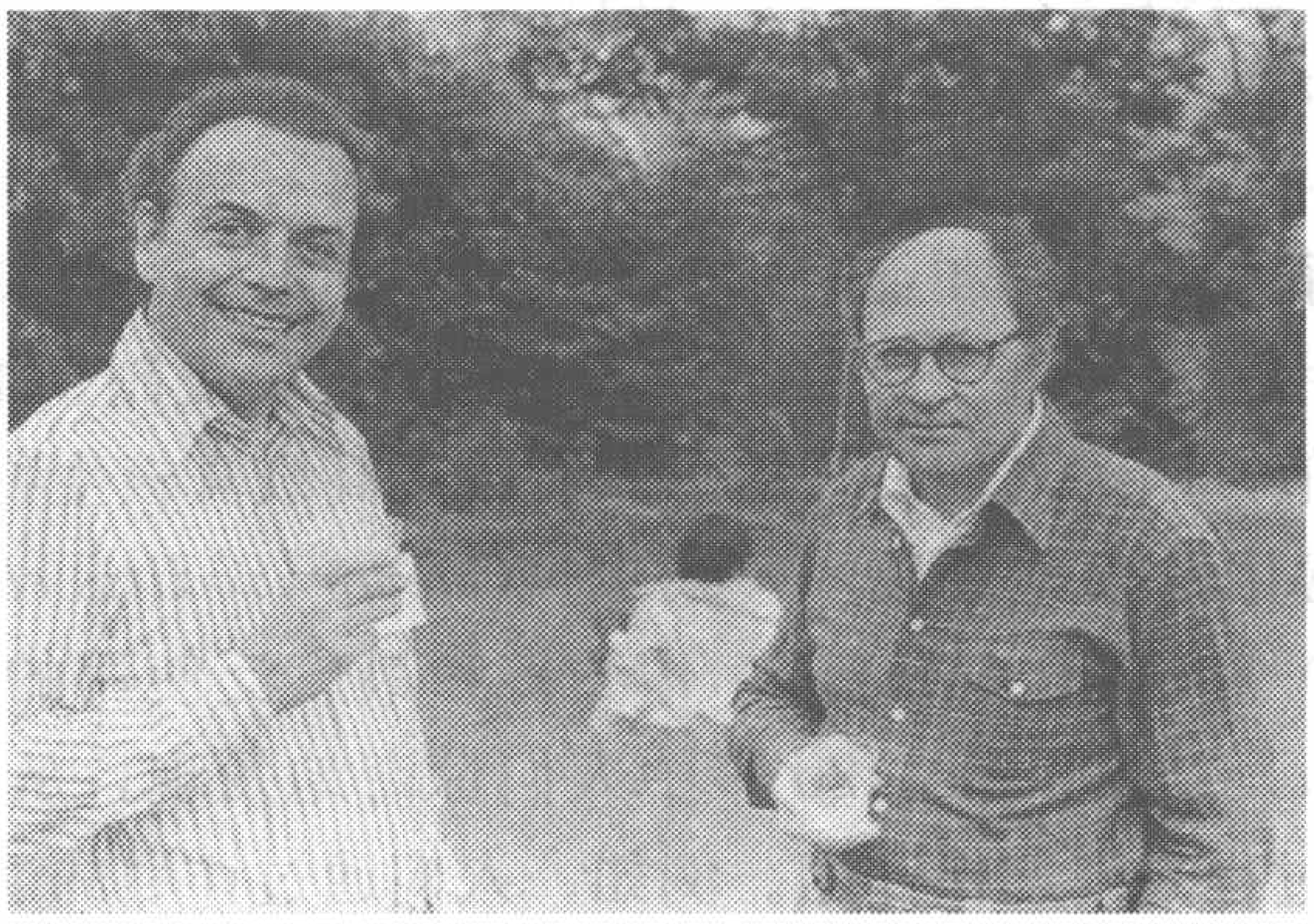
对分子生物学家们来说20世纪70年代是一个不平静的时期。很多人感到原核生物如细菌及其病毒已经得到征服，下一轮生物学问题的解答将有赖于真核细胞实验。最易取得的真核细胞是酿酒酵母，它和细菌相似，生活周期极短，但具备更为复杂的真核细胞的很多特性。人们用酵母开展了大量研究，包括DNA复制、细胞周期和转录调控。这些研究极富科学价值，因为人们发现酵母具有很多高级真核生物所共用的分子机制。

第22章以3种常用的模式动物——线虫（*Caenorhabditis elegans*）、果蝇（*Drosophila melanogaster*）和小鼠（*Mus musculus*）和最广泛使用的模式植物——拟南芥（*Arabidopsis thaliana*）结束。过去20年中人们惊奇地发现，从线虫到人类的很多种动物中，不少遗传学过程是高度保守的。例如，人们通过详尽的遗传筛选在果蝇中鉴别了许多信号传导途径以及调控基本发育过程的基因，后来发现在其他高等动物中也同样具有这些信号传导途径中的基因。在转基因小鼠中所建立的高度复杂的基因操作方法帮助研究者确定了在果蝇中发现的遗传途径所控制的那些过程；基因工程小鼠还可作为检验人类疾病的有关假说和治疗方法的模型，如阿尔茨海默病、帕金森病和风湿性关节炎等。

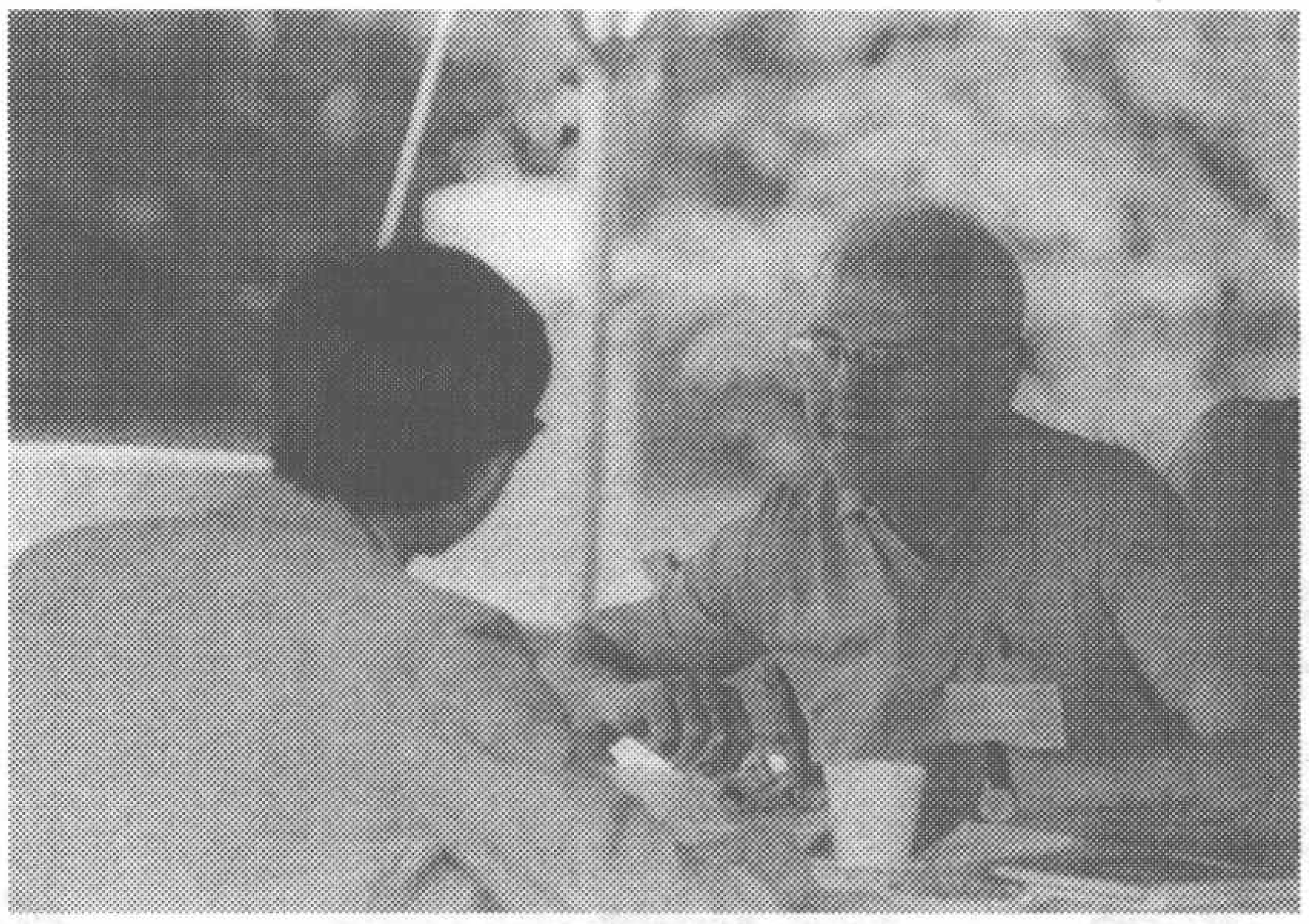
冷泉港实验室档案中的一些照片



Seymour Benzer，1975年突触研讨会。Benzer用噬菌体进行遗传学实验，提出了突变的最小单位的概念，后来发现单个核苷酸即是突变的最小单位（第22章）。他通过功能互补实验还对基因作了实验定义，当时称之为顺反子。后来他的工作主要是用果蝇作模式生物研究行为学。



Werner Arber 和 Daniel Nathans，1978 年 DNA 复制和重组研讨会。 由于发现了限制性内切核酸酶及其在 DNA 分子分析中的应用（第 21 章），他们与 Hamilton Smith 一起，荣获了 1978 年的诺贝尔生理学或医学奖。这一重大发现对 20 世纪 70 年代早期重组 DNA 技术的建立起了关键作用。



Dale Kaiser，1985 年发育的分子生物学研讨会。 Kaiser 在 λ 噬菌体繁殖的早期研究中做出了重大贡献（第 16 章）。他的部分工作使他意识到含有互补单链末端的 DNA 分子可以很容易地连接在一起，这对重组 DNA 技术的建立起了重要作用。



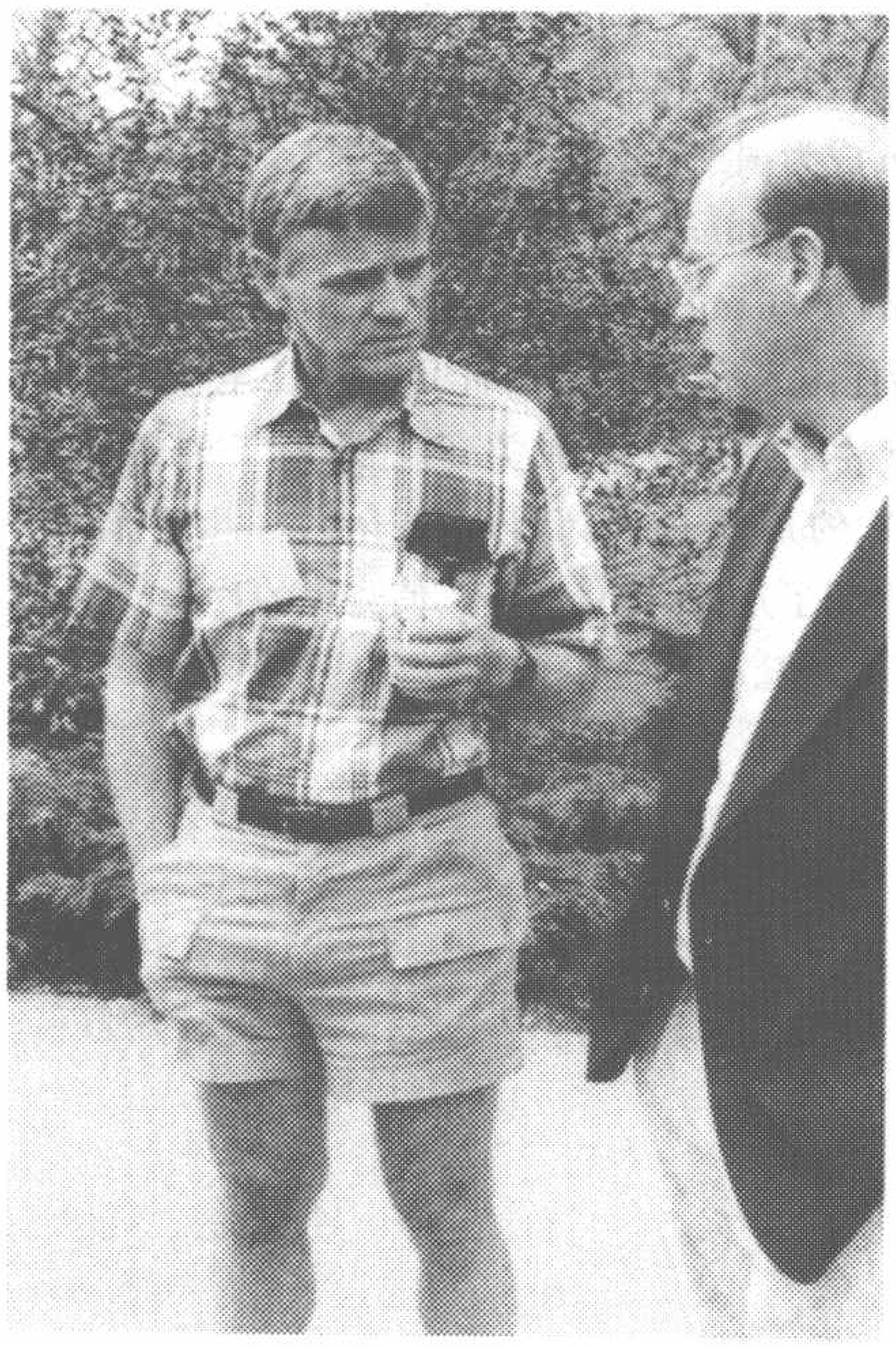
Paul Berg，1963 年大分子的合成和结构研讨会。Berg 是体外构建重组 DNA 分子的一个先驱者，是 1980 年诺贝尔化学奖获得者之一。



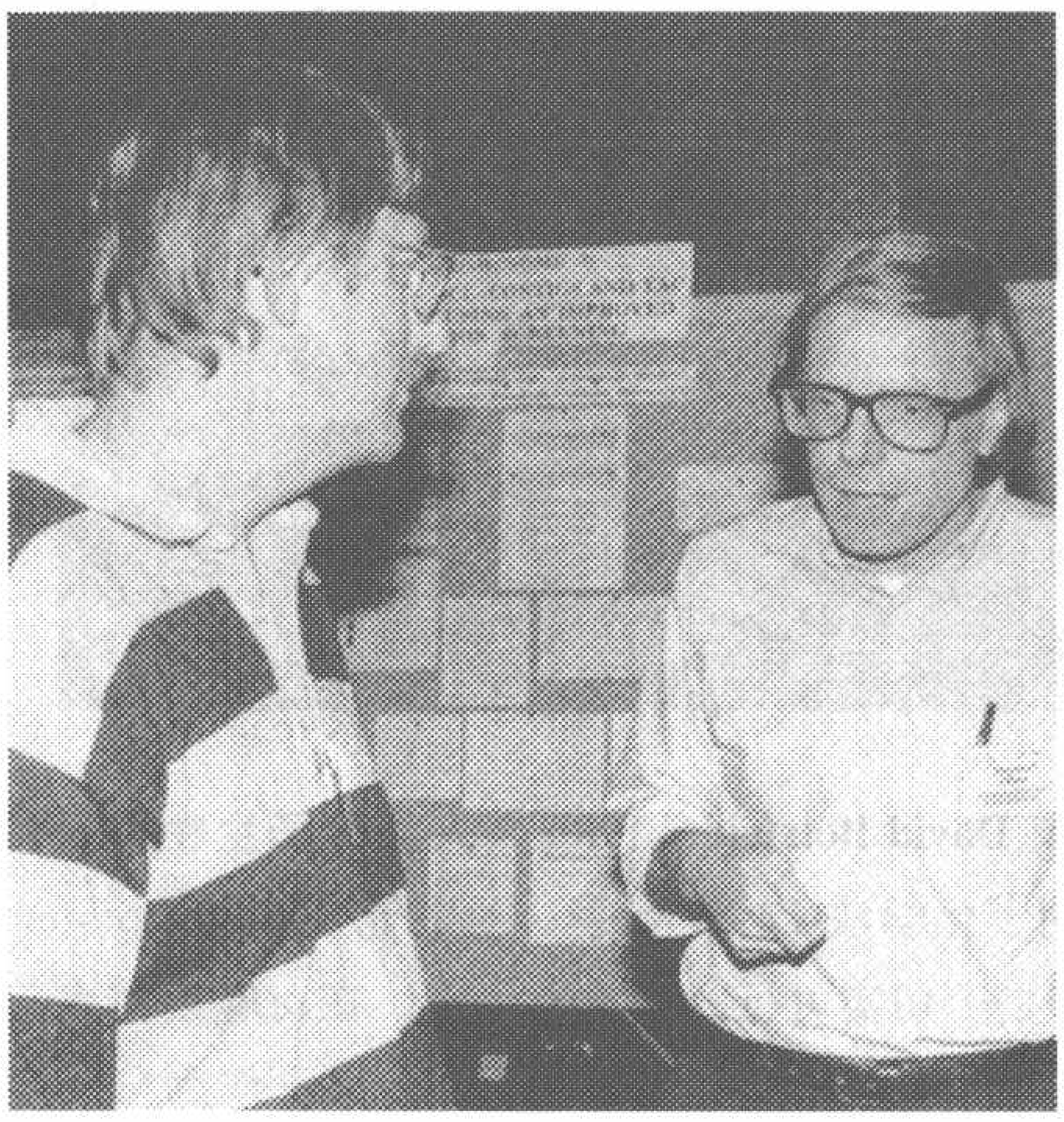
Walter Gilbert 和 David Botstein，1986 年人类分子生物学研讨会。Gilbert 发明了化学法 DNA 测序。图中他正在与 Botstein 进行人类基因组是否切实可行的历史性争论。在对噬菌体多年研究的基础上，Botstein 对使用酵母 *S. cerevisiae* 作为分子生物学研究的模式真核生物的工作做出了重大贡献；他也是基因组学领域发展早期的一个关键人物（第 20 章和第 22 章）。



Albert Keston，Sidney Udenfriend 和 Frederick Sanger，1949 年氨基酸和蛋白质研讨会。Keston 发明了检测葡萄糖的测试带，Udenfriend 进行了抗疟疾药物的筛选和实验。Sanger 是唯一两次获得诺贝尔化学奖的人。他第一次获奖是在 1958 年，当时他发明了一种测定蛋白质氨基酸序列的方法；22 年后他获得了第二次诺贝尔化学奖，这一次获奖是因为他发明了 DNA 测序方法，现在人们进行 DNA 测序所使用的几乎都是这种方法，包括用来进行全基因组测序的自动测序仪所使用的测序方法（第 20 章和第 21 章）。“蛋白质具有特定的序列”的发现不仅是显著的技术成就，而且首次揭示了蛋白质可能也有特定的结构。



Leroy Hood 和 J. Craig Venter，1990 年基因组作图和测序会议。Hood 发明了自动测序技术，并在 20 世纪 80 年代中期制造了第一台仪器。Venter 最早使用改进了这种自动测序仪，得益最大。依靠对原始数据的拼接，这种仪器可实现鸟枪法测序策略，由此极大地加速了全基因组的测序工作，其中包括人类基因组测序（第 20 和 21 章）。



Francis S. Collins 和 Maynard V. Olson，1992 年基因组作图和测序会议。 Collins 是最早的“基因狩猎者”之一，他在 1989 年就发现了广为关注的囊性纤维化基因。1993 年，他接替 James Watson 成为国家人类基因组研究中心（National Center for Human Genome Research，National Human Genome Research Institute 的前身）的主任，后任国家人类基因组研究所主任至今。在这幅照片上，他正站在关于酵母人工染色体（YAC）的展板前听 Olson 的介绍，几年前，Olson发明了这种 YAC 载体，使克隆的片段长度增加了 10 倍。



Eric S. Lander（发言人）1986 年人类分子生物学大会。 Lander 后来成为公众资助的人类基因组计划（Human Genome Project）的领军人物，他也是 2001 年发表的基因组序列文章的第一作者（第 20 章）。就像在前面关于 Gilbert 和 Botstein 的照片一样，Lander 正在 1986 年的辩论中就是否有必要进行人类基因组测序发表他的观点。在他旁边是 David Page，看上去若有所思，他的工作主要围绕 Y 染色体的结构、功能和进化，而前排的 Nancy Hopkins（一名发育生物学家，也是本书第四版的作者之一）和后排的 James Watson 则听得津津有味。

第21章 分子生物学技术

本章概要

- 核酸
- 蛋白质
- 蛋白质组学
- 核酸-蛋白质相互作用

我们所看到的活细胞是一个十分复杂的实体，能产生成千上万种大分子，并含有一个从几百万到几十亿碱基对的基因组。理解细胞的遗传过程是如何进行的需要多种多样具挑战性的实验方法，这包括第 22 章中将要讨论的合适的模式生物的使用，而模式生物使遗传分析手段可以得到使用，也包括这一章我们要讨论的从细胞混合物中分离单个大分子的方法，以及将基因组分割成可处理的片段以便于特定DNA 序列的操作和分析的方法。在过去的几十年里，上述实验方法的成功建立是分子生物学发展的主要推动力之一，也是分子生物学所取得的重要成就。

近来，应用分子生物学方法进行细胞中全部 RNA 和蛋白质的大规模分析并测定完整基因组的核苷酸序列已经成为可能。这些基因组和蛋白质组研究方法，加上迅速增加的、已测得基因组序列，使不同生物基因组的大规模比对以及在一个特定的细胞类型中鉴定所有的磷酸化蛋白成为可能。

本章我们将简要介绍这些分子生物学方法、基因组学方法、蛋白质组学方法，以及它们的基本原理。我们将会看到，这些分子生物学方法的建立依赖于对生物大分子自身特性的理解。例如，对 DNA、RNA 结构和碱基互补特征的了解促进了杂交和测序技术的建立，使基因结构和基因表达的快速详尽分析成为可能；对 DNA 聚合酶、限制性内切核酸酶和 DNA 连接酶的深入研究诞生了 DNA 克隆技术和聚合酶链反应（PCR），这些技术使科学家们可以分离大量的任意 DNA 片段，甚至能从已灭绝的生命体中分离DNA 片段。

结构和功能的多样性使蛋白质分析的难度更大，在很多情况下每一个蛋白质都有其特殊性。不过，科学家们也发展了多种用于蛋白质纯化和功能鉴定的技术。近年来，质谱技术的应用使人们在鉴定混合物中的微量蛋白质，甚至鉴定蛋白质中氨基酸的特定修饰（如磷酸化）能力有了很大的提升。这些技术使研究者能够在总体水平研究蛋白质群体，而不再局限于原有的逐一鉴定的方式。

本章分为四个部分。第一部分主要讲述核酸的操作和分析技术，从 DNA、RNA 的分离到完整基因组的测序和比较基因组学。接下来的两部分主要介绍蛋白质的分离和分

析：单个蛋白质的纯化，以及运用蛋白质组学方法分析细胞或者组织中的全部蛋白质。虽然从细节上看，这些技术之间并不相似，但我们可以看出，分离和操作核酸和蛋白质的许多程序都是基于相同的原理。本章的第四部分主要为大家介绍核酸-蛋白质相互作用的分析方法，以及探索那些独立的成分之间的聚集、相互作用，最终构成有活力的细胞的方法。

最后，请注意在接下来的讨论中，当谈及一种特定大分子的分离和纯化时，我们并不是指分离单个分子而是指将细胞中大量的某种分子与所有其他分子分离开来。

核酸

凝胶电泳能根据大小分离 DNA 和 RNA 分子

我们先讨论用凝胶电泳（gel electrophoresis）技术分离 DNA 和 RNA 分子。在外加电场的作用下，线性 DNA 分子在一种称为凝胶基质（gel matrix）的果冻样惰性多孔物质中，能按大小分开。由于 DNA 带负电荷，在图示的外加电场作用下，它们在凝胶中向正极迁移（图 21-1）。DNA 分子是柔性的，占据一定的有效体积。凝胶基质中的网孔能按照 DNA 的体积对它们进行分筛。由于大分子的有效体积比小分子大，较难通过凝胶孔隙，在凝胶中迁移得就比小分子慢。这就意味着，一旦凝胶“跑”一定时间后，不同大小的分子由于在凝胶中迁移距离的不同而被分离开来。

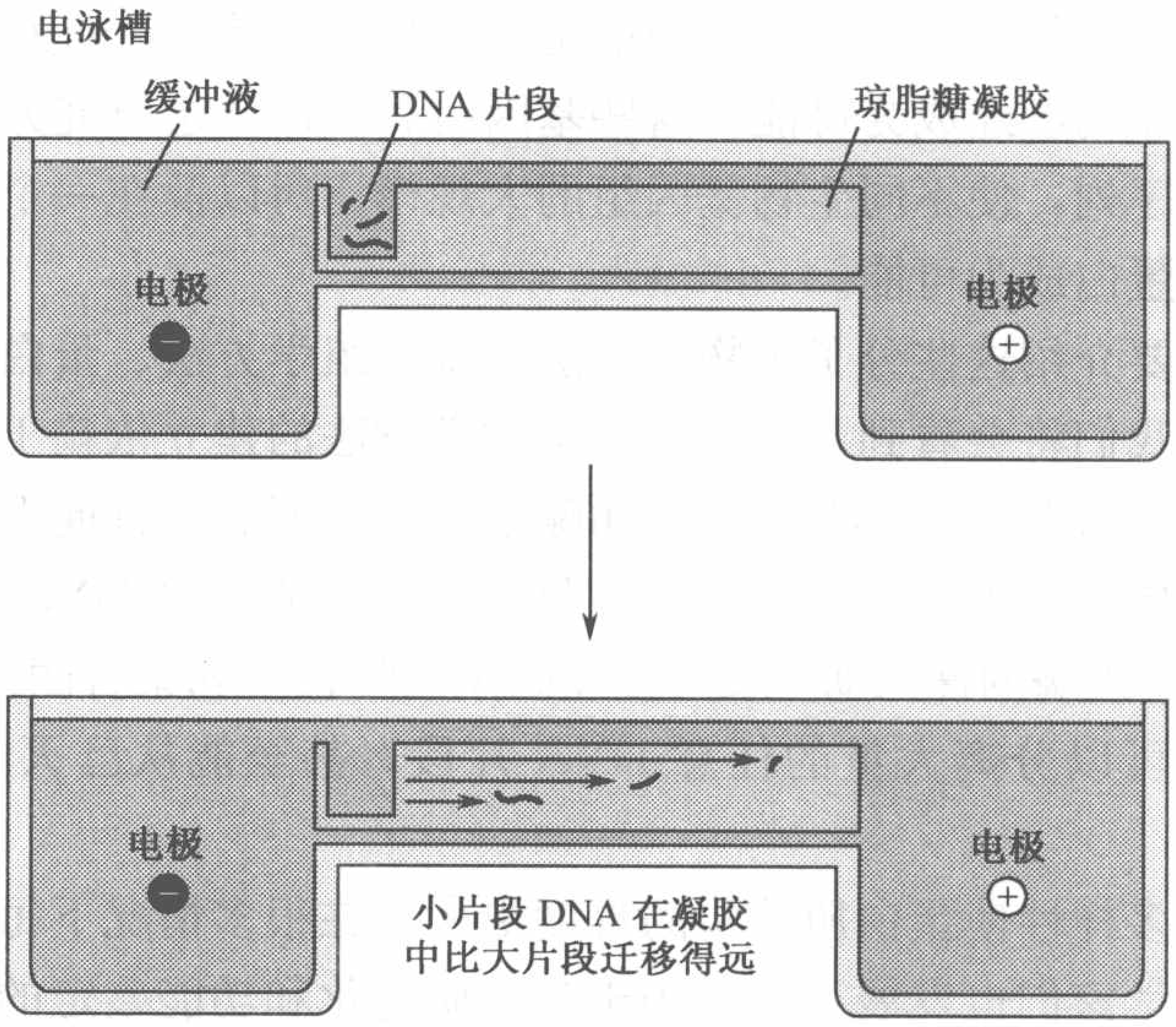


图 21-1 凝胶电泳分离 DNA。图中所示为一块胶的侧切面。图中左侧是 DNA 混合物的上样孔，算是胶的顶部，也是电场的阳极端，胶的底部是阴极。带负电荷的 DNA 片段将从胶的顶部向底部移动。如图所示，DNA 片段的迁移距离和其大小呈负相关。（经许可，仿自 Micklos D. A. and Freyer G. A. 2003. *DNA science*：*A first course*，2nd ed.，p. 114. © Cold Spring Harbor Laboratory Press.）

在电泳完成后，用荧光染料如溴化乙锭（ethidium bromide）对凝胶进行染色，可以显示DNA分子。荧光染料能和DNA分子结合，嵌入到堆积的碱基中（图6-28）。每一条带反映了一群特定大小的DNA分子的存在。

通常使用的凝胶基质有两种：聚丙烯酰胺凝胶（polyacrylamide）和琼脂糖凝胶（agarose）。聚丙烯酰胺分辨力高，但能区分DNA的大小范围窄。因此，聚丙烯酰胺凝胶电泳能区分只相差一个碱基对的DNA片段，但仅限于几百碱基对（低于1000）大小的DNA片段。琼脂糖凝胶分辨力不如聚丙烯酰胺凝胶，但能区分几十至几百kb（千碱基对）大小的DNA片段。

极长的DNA片段即使在琼脂糖凝胶中也不能穿过其网孔。它们在凝胶基质中蛇行，一端在前，另一端在后。因而，超过一定大小(30～50kb）的DNA分子迁移程度相似而难以区分。但是，如果采用电流方向相互垂直的脉冲式电场，这些极长的DNA片段就能区分开来。这种技术就是脉冲场（pulsed-field）凝胶电泳（图21-2)。每次电场方向改变的时候，在凝胶中蛇行的DNA分子就随之调整其泳动方向。DNA片段越大，其调整泳动方向所需的时间就越长。脉冲场凝胶电泳可以用来测定整套细菌染色体和低等真核生物如真菌染色体的大小，即长达数个Mb的DNA分子。

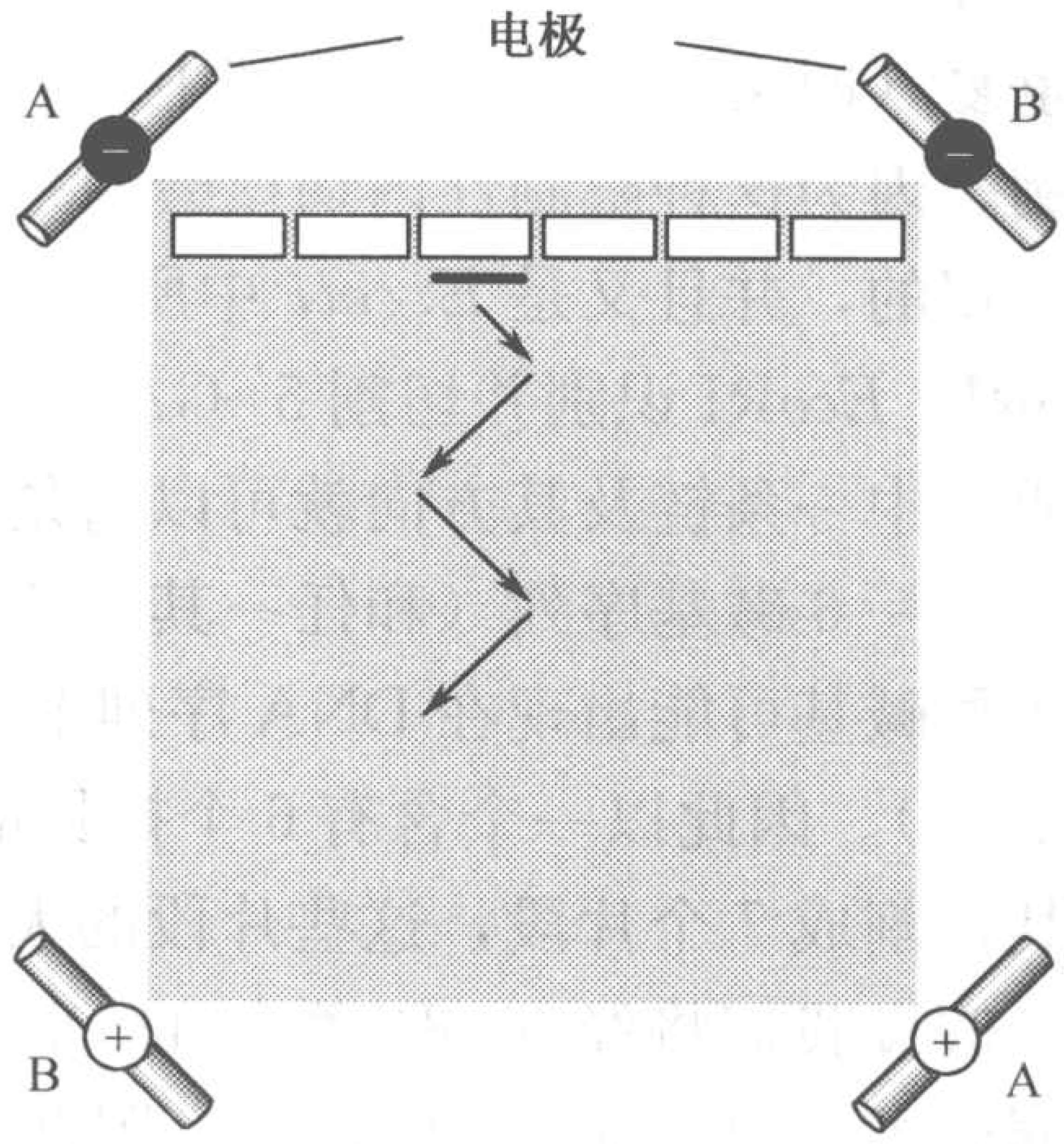


图21-2 脉冲场凝胶电泳。本图是凝胶的俯视图，凝胶顶部有一系列样品孔。A、B代表两套电极，如文中所述，它们交替开关。A组电极通电时，DNA向凝胶的右下角泳动，即A组电极的阳性末端。当A组电极断电而B组电极通电时，DNA向凝胶的左下角泳动。图中箭头标示了电泳进行时DNA的迁移路径。(经许可，仿自Sambrook J. and Russell D. W. 2001. *Molecular Cloning*: *A laboratory manual*, 3rd ed., Fig. 5-7. © Cold Spring Harbor Laboratory Press.)

电泳不仅是根据DNA分子的分子质量大小来分离它们，还按照它们的形状和拓扑学特性对它们进行分离（图6-26)。松弛型的或带缺口的环状DNA分子比相同质量的线性分子迁移得慢。再者，如我们所看过的，超螺旋DNA由于致密而有效体积小，电泳时比相同质量的超螺旋程度较轻或松弛型环状DNA分子迁移速度快。

电泳也能用来分离RNA。线性双链DNA分子二级结构均一，它们电泳时的迁移率和其分子质量大小成比例。和DNA相似，RNA也是均一的负电荷。但是RNA分子通常是单链的，并含有大量二级和三级结构（第6章），这些都会影响它们的电泳迁移率。为了解决这个问题，RNA分子可以先用乙二醛等试剂处理，这些试剂能和RNA分子反应而防止碱基对的形成（乙二醛能和碱基中的氨基基团形成加成化合物，从而防止了碱基配对）。乙醛酸化的RNA分子不能形成二级或三级结构，因此能以和分子质量大致成比例的迁移率迁移。后面我们可以知道，电泳也可以以相似的方式将蛋白质按大小分离。

限制性内切核酸酶在特定位点切割 DNA 分子

大多数天然 DNA 分子因太大而不适于在实验室中操作或分析。就我们所知，染色体是含有上千个基因的具有超过 1 亿碱基对的单个 DNA 分子（第 7 章）。如果我们要研究 DNA 上的个别基因和位点，必须先将细胞中的大 DNA 分子分解成可操作的片段。这可以用限制性内切核酸酶（restriction endonuclease）来完成。限制性内切核酸酶是一类识别特定的 DNA 序列并在特定位点切割 DNA 的核酸水解酶。

分子生物学中使用的限制性内切核酸酶通常识别 4～8bp 的短的靶序列，一般为反转重复（回文）序列，并在其中的特定位置切割。以常用的限制性内切核酸酶 *Eco*RI 为例。因为这种限制性内切核酸酶是在一些特定的大肠杆菌（*Escherichia coli*）菌株中被发现的，并且又是 *E. coli* 中所发现的第一个限制性内切核酸酶，人们就将它命名为 *Eco*RI。*Eco*RI 识别并切割 5′-GAATTC-3′。（由于 DNA 的两条链是互补的，我们只需说明其中一条链及其方向就可以清楚地描述识别序列。）

这一 6 碱基序列（和任一其他 6 碱基序列一样）估计平均每 4kb 出现一次（这是因为 4 种碱基可能出现在 DNA 序列上的任一位置，于是任一特定 6 碱基序列出现的概率为 $1/4^6$）。因此以一个含有 6 个拷贝的 GAATTC 序列的线性 DNA 分子为例：*Eco*RI 能将其切割成 7 个片段，这些片段的大小反映了这个 DNA 分子中这类酶切位点的分布情况。假设我们将经 *Eco*RI 酶切的 DNA 在凝胶中进行电泳：这 7 个片段会按它们大小的不同而被分离开来（图 21-3）。因此，在所示的实验中，*Eco*RI 已将 DNA 切割成特定片段，每一片段都对应于 DNA 分子上的特定区域。

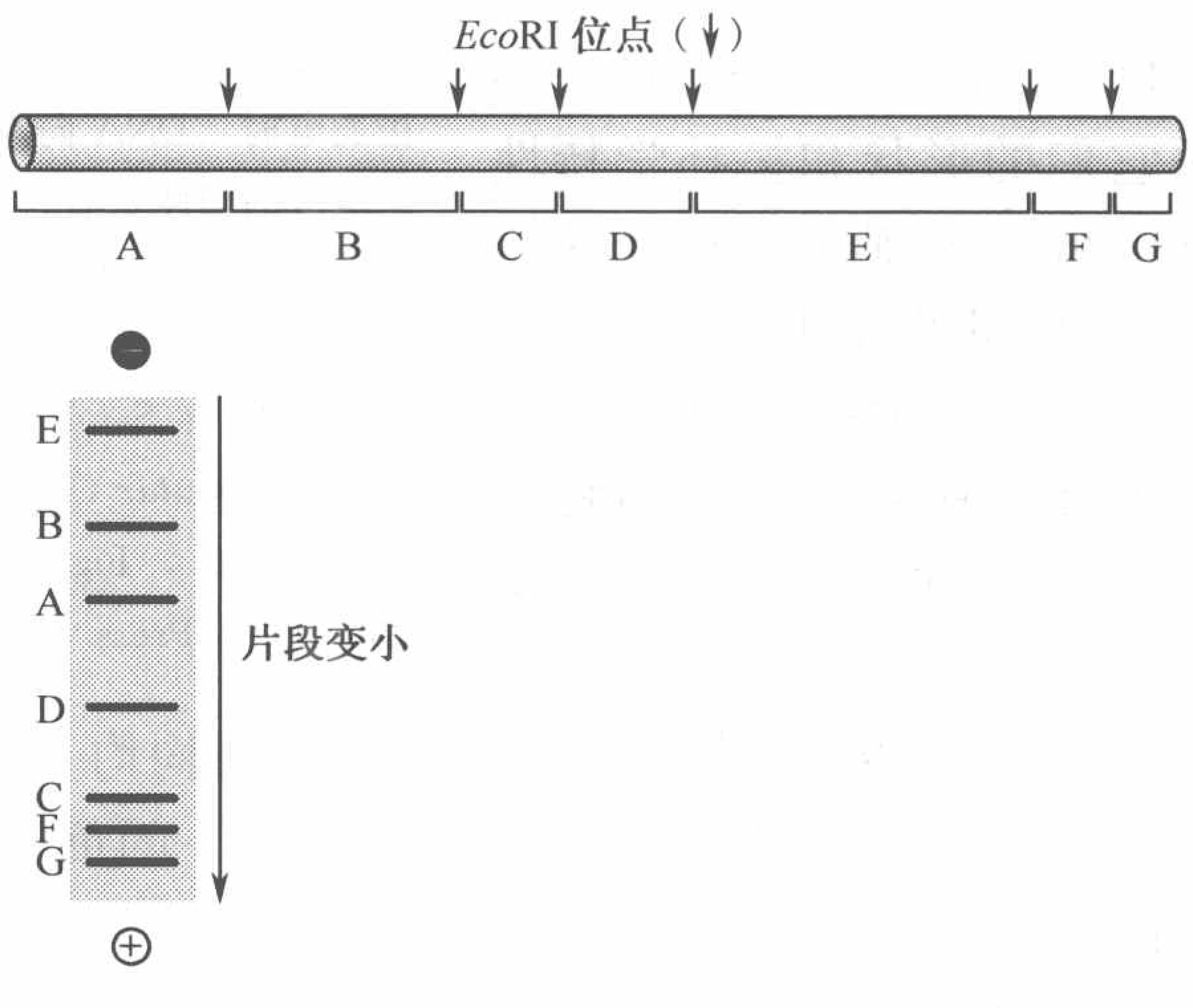


图 21-3 限制性内切核酸酶 *Eco*RI 消化某一 DNA 片段。上部所示为一个 DNA 分子和其中 *Eco*RI 的酶切位点。DNA 分子用 *Eco*RI 消化后，在琼脂糖凝胶中电泳，可观察到图中所示的带型。

如果将同一 DNA 分子用另一种限制酶裂解，如 *Hind*Ⅲ（它也识别 6 个碱基的靶序列，不过这个序列为 5′-AAGCTT-3′），这个 DNA 分子就会在另外一些位点上被切

割而产生其他大小不同的片段。因此，使用多个限制酶可以分离一个 DNA 分子的不同区域，也能用来鉴定特定的 DNA 分子。一个特定的 DNA 分子用一组不同的限制酶消化时，能产生一系列特征性的片段。

其他的限制酶，如 *Sau*3A1（发现于金黄色葡萄球菌（*Staphylococcus aureus*）中）识别 4 碱基序列（5′-GATC-3′），因此能更频繁地切割 DNA，大约每 250bp 切割一次。另外一个极端是 *Not*I，它识别 8 碱基序列（5′-GCGGCCGC-3′），平均每 65kb 才切割一次（表 21-1）。

表 21-1　一些限制性内切核酸酶和它们的识别序列

酶	序列	切割频率[a]
*Sau*3A1	5′-GATC-3′	0.25kb
*Eco*RI	5′-GAATTC-3′	4kb
*Not*I	5′-GCGGCCGC-3′	65kb

a 频率＝1/4n，n 为所识别序列的碱基对数目。

需要注意的是，有些限制性内切核酸酶是对甲基化敏感的，也就是说，限制性内切核酸酶识别位点中核酸的甲基化，会抑制限制性内切核酸酶在该位点的酶切作用。限制性内切核酸酶不仅识别序列的特异性和长度有所不同，而且它们所产生的 DNA 末端的性质也不相同。一些酶如 *Hpa*I 产生的是平末端，其他酶如 *Eco*RI、*Hind*Ⅲ和 *Pst*I 产生的是交错末端（图 21-4）。例如，*Eco*RI 切割两条链的交错位置上的 G 和 A 之间的共价键（磷酸二酯键）。这两个切割位点之间的 4 个碱基对间的氢键容易断裂，而产生 5′多出 4 个核苷酸的末端（图 21-5），注意这些末端彼此互补。因为这些末端之间或与用同一种酶切割的其他 DNA 分子之间能通过碱基配对而退火，我们通常将它们称为“黏

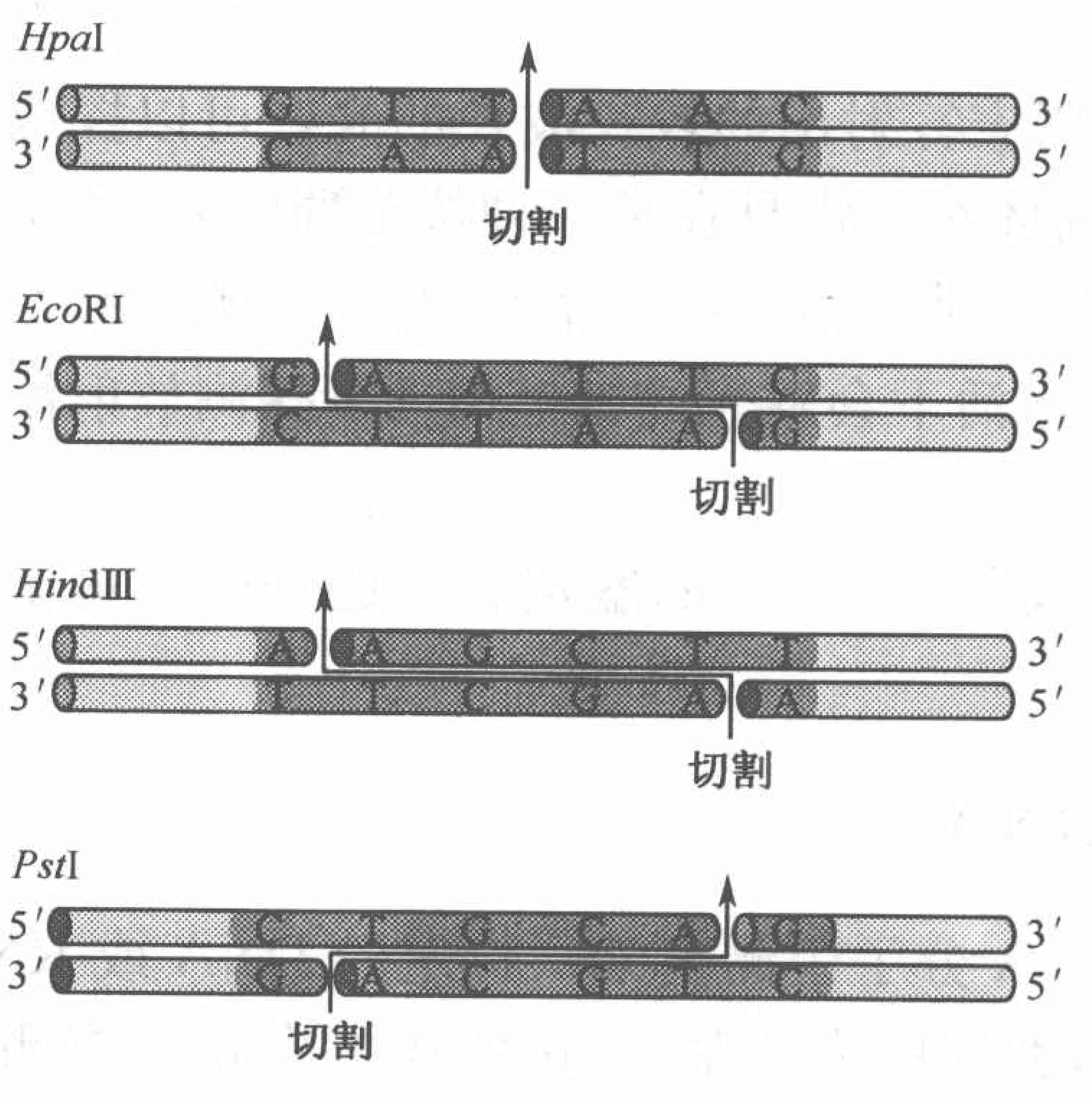


图 21-4　不同限制酶的识别序列和切割位点。如图所示，不同限制性内切核酸酶不仅识别不同的靶位点，而且也在这些位点中的不同位置做切割。因此能产生平末端或 5′、3′突出的末端。

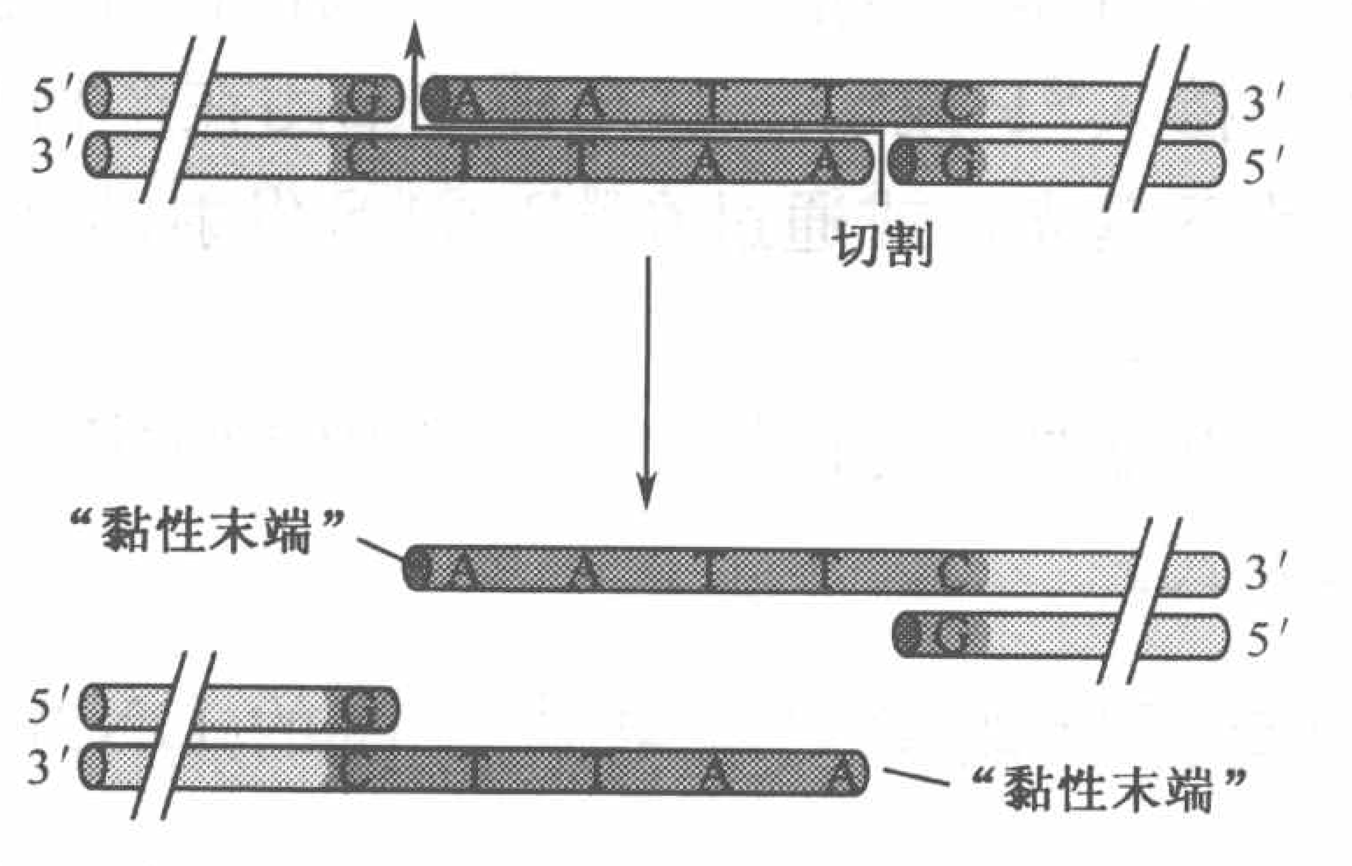


图 21-5　一个 *Eco*RI 位点的切割。*Eco*RI 在两条链的识别位点内切割，从而产生 5′突出末端。因为这些末端与用同一种酶切割的其他 DNA 分子之间能通过碱基配对互补而黏合，我们通常将它们称为“黏性末端”。

性末端”。这一特性在我们讨论 DNA 克隆时很有用。

DNA 杂交能用来鉴定特定的 DNA 分子

如同我们在第 6 章所述，变性 DNA 重新退火（互补链间重新形成碱基对）的能力使来源不同但具有一定同源性的变性 DNA 在适宜的离子强度和温度下混合时可以形成杂交双链。两个不同来源的互补单链多聚核苷酸之间的碱基配对过程叫做杂交(hybridization)。

许多技术依赖于两个序列互补的 DNA 分子之间杂交的特异性。例如，这一特性是鉴定复杂核酸混合物中特定序列的基础。在这种情况下，其中一个分子是特定序列的一个探针（probe）——一个经纯化的片段或化学合成的 DNA 分子。这个探针用来在核酸混合物中寻找含有与其互补的序列的分子。探针 DNA 必须先被标记，这样一旦它发现靶序列后，就可以轻易将其定位。被探针标记的混合物通常已经在凝胶上按大小做分离，或者作为一个文库而分布在不同菌落中（见下文）。

标记 DNA 有两种基本方法。第一种方法是在一个被标记的前体物存在的情况下，合成新的 DNA 分子，对此我们将在后面加以描述。另一种方法是在一个完整的 DNA 分子末端加上一个标签。例如，多聚核苷酸激酶会将 ATP 的 γ-磷酸基团加到 DNA 的 5′-OH 末端。如果这个磷酸基团具有放射活性，磷酸基团的转移能使其 DNA 分子受体被标记上放射活性。

掺入标记法（另一种机制）通常是通过一个如下文所描述的被标记的前体物进行聚合酶链反应（PCR)，或者通过把随机的短小的六聚核苷酸杂交到 DNA 上以使 DNA 聚合酶能延伸它们。被标记的前体物通常是用荧光染料或放射活性物质修饰的核苷酸。通常荧光染料只需和用作 DNA 合成前体物质的 4 种核苷酸的任一碱基混合（对大多数前体而言，25％的标记一般就可以满足需要）。

用适宜波长的紫外线照射 DNA 样品，检测应答射出的较长波长的光，可以探测到用荧光前体物标记的 DNA。放射活性标记的前体物通常是在某一种核苷酸的 α-磷酸基团中掺入 ^{32}P 或 ^{35}S。这个磷酸基团在产物 DNA 中被保留（第 8 章)。将需检测样品在 X 光片下曝光，或通过在 ^{32}P 或 ^{35}S 发射的 β 粒子激活下会发光的光电倍增管探测放射活性的 DNA。

杂交技术可以以多种方式用于特定 DNA 或 RNA 片段的鉴定。下述是两种最常见的方式。

杂交探针能鉴定经电泳分离的 DNA 和 RNA

这种方法常用来在一群相似的 DNA 和 RNA 分子中测定某一特定的 DNA 或 RNA 分子的丰度或大小。例如，测定两种不同细胞类型中表达某一 mRNA 的含量，或测定一个含有你所要研究的基因的限制性片段的长度。通过使用印迹方法，定位特定的经电泳分离的核酸，可以得到这类信息。

假设你用 *Eco*RI 酶切酵母基因组后，想知道含有目的基因的限制性片段的长度。但用溴化乙锭染色后，切割酵母基因组所产生的上千个 DNA 片段在凝胶上 4kb 左右形

成连续分布的条带，我们无法将如此多的片段区分成相互分散的带型。Southern 印迹杂交技术（因其发明者 Edward Southern 而得名）可以在连续分布的条带中鉴定含有目的基因的限制性片段的长度。

切割所得的 DNA 经凝胶电泳分离后，先在碱性溶液中浸泡，使双链 DNA 片段变性。然后我们将这些片段转移到带正电荷的膜上，这些片段结合于膜后产生一个印迹。在这一转移过程，DNA 片段在膜上的位置对应于它们在凝胶电泳中迁移的位置。目的 DNA 被转移到膜上后，负荷的膜与非特异性的 DNA 片段混合物共同孵化以使膜上剩余的结合位点达到饱和。因为混合物中的 DNA 随机地分散在膜上，而通过恰当的筛选将不包括目的序列，所以不会干扰下一步骤的目的序列检测。

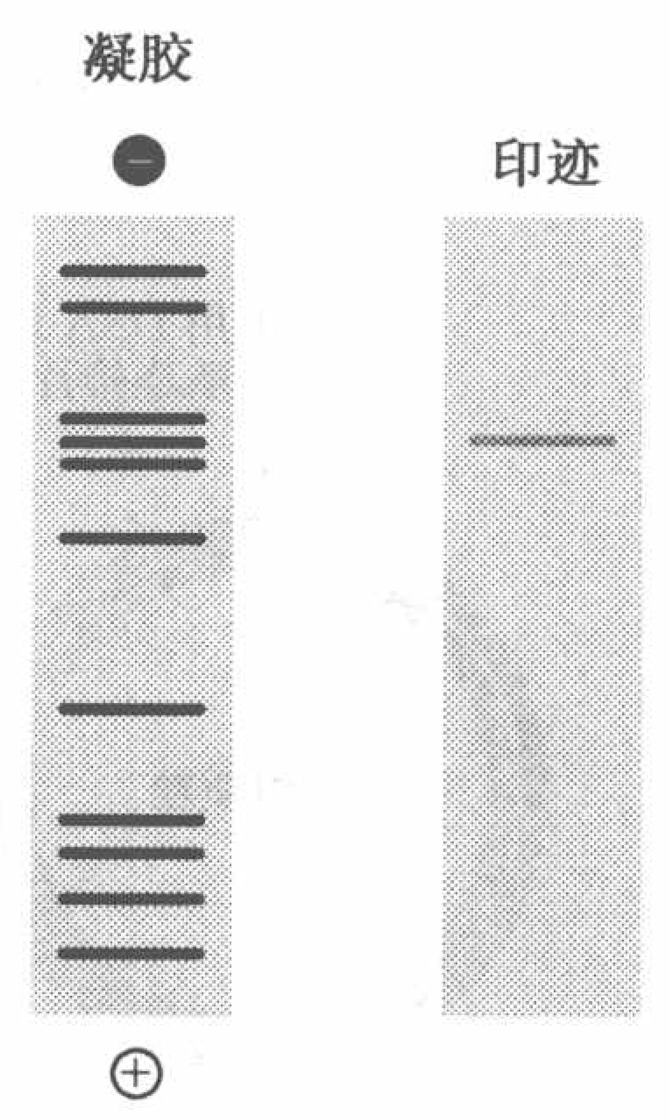


图 21-6 Southern 印迹。DNA 分子经限制性内切核酸酶消化产生的 DNA 片段在琼脂糖凝胶中电泳。染色后，可以看到一系列条带。将电泳条带转移至滤膜，用只和 DNA 分子的一个特定序列同源的 DNA 片段作探针进行杂交，能看见单一条带，其位置对应于含有此序列的片段在凝胶上的位置。

然后结合于膜的 DNA 片段与探针 DNA 一起进行孵化。探针 DNA 含有和目的基因互补的一段序列。因膜上所有非特异结合位点已被占据，探针 DNA 只有与在膜上与其互补的 DNA 杂交才能结合到膜上。这个过程所需的离子浓度和温度与核酸变性、复性的条件相似。在严格控制的离子浓度和温度下，探针 DNA 只与它的互补序列紧密结合。探针的量通常远远超过固定于膜上的靶片段，这有利于杂交反应的进行，而非变性 DNA 的重新退火。变性 DNA 在滤膜上的固定也能干扰其重新退火。印迹上探针杂交的位置可以用对标记 DNA 发出的光或电子敏感的胶片或其他介质检测到。例如，将 X 光片曝光于滤膜，然后显影，所得的放射自显影图像（autoradiogram）和滤膜上的杂交分子位置对应（图 21-6）。

在 RNA 群中鉴定某一 mRNA 的步骤与此相似，但叫做 Northern 印迹杂交（以示区分）。由于 mRNA 比较短（常小于 5kb），它们无需再用任何酶消化（特异性的 RNA 裂解酶也不多）。其他步骤都和 Southern 印迹大同小异。分离的 mRNA 转移至带正电荷的膜上，和放射性标记的探针反应。（此处是 RNA 和 DNA 之间的互补链碱基配对形成杂交体。）

Northern 印迹杂交更适于确定一个样品中特定 mRNA 的含量，而不是其大小。特定 mRNA 的含量能反映编码它的基因的表达水平。因此，使用目的基因诱导剂诱导一个细胞后，可以用 Northern 印迹杂交检测与未经诱导的细胞相比某一 mRNA 水平的升高。Northern 印迹杂交也可以用来比较同一生物体不同组织间特定转录物的相对水平（也就是目的基因的表达水平）。由于在这些分析中我们使用了过量的 DNA 探针，杂交产物的量与原始样品中的 mRNA 量相关，从而可以测定 mRNA 的相对量。

Southern 印迹杂交和 Northern 印迹杂交的原理也是第 20 章中讨论的基因芯片分析的基础。大量序列信息的获得使“反向杂交”实验的发展成为可能。一张芯片的制作需要运用几百至几千个已知的 DNA 序列。其中每个序列对应所研究的生物体中的不同

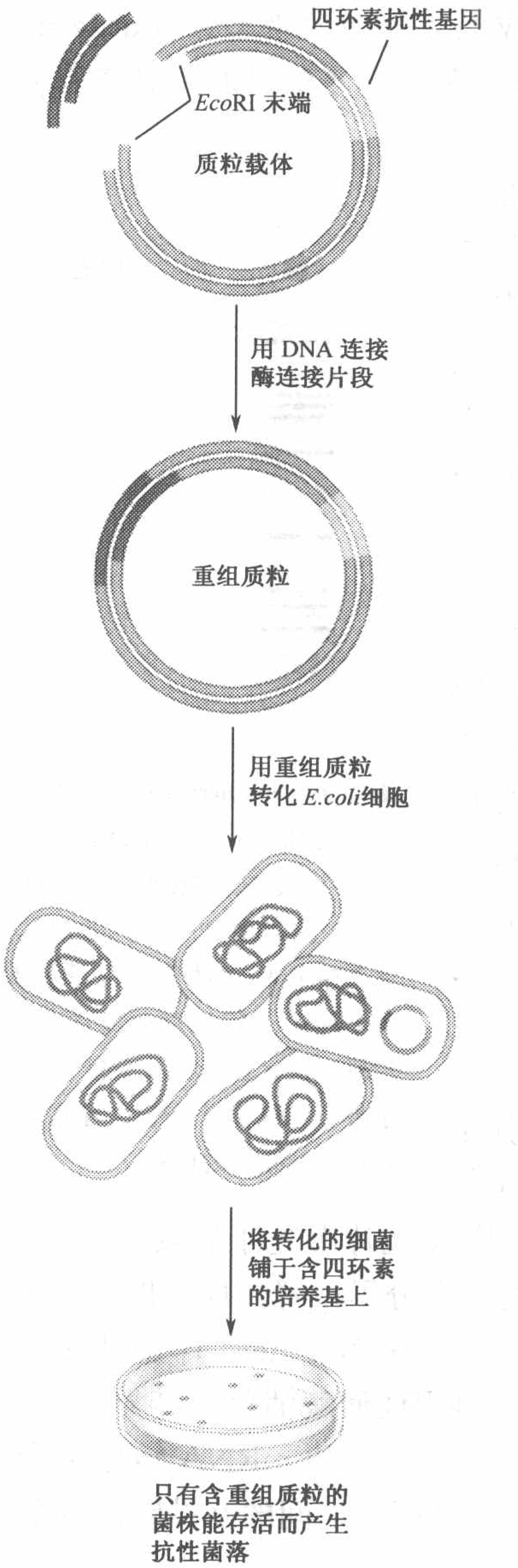


图 21-7 用质粒载体进行克隆。一个用 *Eco*RI 酶切得到的 DNA 片段被插入到用同一种酶切所得的线性质粒载体。连接后，重组质粒转化细菌（见正文）。由于质粒有抗生素抗性，带有质粒的细胞能在含有抗生素的培养基上生长从而加以选择。（经许可，仿自 Micklos D. A. and Freyer G. A. 2003. *DNA Science: A first course*, 2nd ed.，p. 129，© Cold Spring Harbor Laboratory Press.）

基因，并被独立固定在芯片阵列中（图 19-1）。请注意这里所用的技术反向运用了 Southern 或 Northern 分析方法。在芯片分析中，固定的、未被标记的序列被称作探针。这些探针是已知的序列。而目的序列则由从细胞或组织中的全体 RNA 生成的 cDNA 经扩增和标记组成的。当目的序列与阵列中的 DNA 探针杂交时，芯片中每个 DNA 杂交信号的强弱反映了目的基因的表达水平。

特定 DNA 区段的分离

为了对基因及其功能做分子生物学分析，我们需要先将较大的 DNA 分子分解成一些特定的 DNA 区段并做选择性扩增。这样我们就能够分析这个特定 DNA 分子所编码的信息。从而，这段 DNA 可以被测序分析，或可以被克隆，表达以研究其蛋白质产物。

大量纯化特定 DNA 分子的能力也使我们能用各种其他方式操作它。比如，我们就能制备重组 DNA 分子。重组技术可以用来改变某一特定基因的表达（如将其编码序列和其他启动子融合）或产生编码所谓融合蛋白（即各个部分来自于不同蛋白质的杂合蛋白）的 DNA。DNA 克隆和 PCR 扩增技术已成为研究基因表达调控、基因组稳定性和蛋白质功能的基本手段。

DNA 克隆

构建重组 DNA 分子并在细胞中保存这些分子，叫作 DNA 克隆（DNA cloning）。这个过程通常涉及载体（vector）和插入 DNA（insert DNA），载体能提供在细胞中扩增克隆 DNA 的信息，插入 DNA 则包含目的基因并能插入到载体中。制备重组 DNA 分子的关键是在特定序列处切割 DNA 的限制酶以及能连接切割所得 DNA 的其他酶类。通过制备可在宿主中扩增的重组 DNA 分子，我们可以将某一特定的插入 DNA 从其他 DNA 中分离纯化出来并大量扩增。

接下来我们将介绍 DNA 分子的切割、重组和扩增。然后我们还要讨论如何制备大量杂合分子

的集合，即文库（library）。一个文库使用一种通用载体，这个载体携带许多不同的插入片段。除了文库的制备方法外，我们还将讲述如何从文库中鉴定和分离特定的DNA区段。

在质粒载体中克隆DNA

DNA用限制酶切割成片段后，通常需要插入到某一载体中进行扩增。也就是说，为了在宿主中进行复制，DNA片段必须插入到第二个DNA分子（载体）中。最常用于扩增DNA的宿主是大肠杆菌（*E. coli*）。载体DNA通常具备3个特性。

1. 载体含有一个复制起点，它们能独立于宿主染色体进行自主复制。（注意酵母载体还需含有一个着丝粒。）

2. 载体含有筛选标记，这使我们容易鉴别含有载体（和所连DNA）的细胞。

3. 载体含有一个或多个限制酶的酶切位点，从而可以使DNA片段插入于载体的特定位点，而不干扰以上两个功能。

最常用的载体为质粒（plasmid），它们是大小约为3kb的环状DNA分子。这些分子最先源于许多细菌和单细胞真核生物中天然存在的环状DNA分子（第22章）。许多情况下（尽管酵母中不是），质粒DNA常常携带抗生素抗性基因。因此，天然质粒已经具备载体的两个特性：①能在宿主中独立扩增；②携带筛选标记。这些质粒的另一个好处是它们在一个细胞中多拷贝存在，因此增加了能从细胞群中分离到的DNA的量。

有时这些质粒也含有有用的特殊酶切位点。然而，自从这些质粒被发现以来，人们就对它们做了简化和改造，如今一个典型的质粒载体在一小段区域内就有20多个单一酶切位点一般称为多克隆位点。这样我们可以应用多种限制酶对靶DNA进行切割。

将DNA片段插入到载体中比较简单（图21-7）。假设一个质粒载体含有*Eco*RI的特异识别位点。用*Eco*RI处理后质粒变成线性。由于*Eco*RI能产生互补的5′突出末端，这些黏性末端能退火形成带有两个缺刻的环（参考图21-5）。DNA连接酶（DNA ligase）和ATP可以封闭缺刻，重新形成共价闭合环状DNA。靶DNA先用限制酶，在这个例子中是*Eco*RI切割以生成插入DNA。这些插入DNA和过量的用*Eco*RI切割形成杂交用黏性末端的载体DNA混合，并加入DNA连接酶以连接这两个DNA的互补末端。加入与质粒DNA量相比过量的插入DNA可确保大部分载体会和插入DNA连接（图21-7）。

一些载体不仅使特定DNA的分离、纯化成为可能，还能驱动插入DNA内基因的表达。这些质粒叫作表达载体（expression vector），它们在插入位点附近含有从宿主细胞中得到的转录启动子。如果一个基因的编码区（不含启动子）置于插入位点的恰当位置，插入基因就能由宿主细胞转录成mRNA并翻译成蛋白质。人们通常使用表达载体来表达异源基因或突变基因以检测其功能。表达载体也能用来大量生产某一种蛋白质以供纯化之用。而且，人们可以通过选择表达载体中的启动子来方便地对插入片段的表达实行调控。例如，在培养基中加入一种简单的化合物（如一种糖或氨基酸）（见第16章中原核生物转录调控的讨论）。如果基因产物有毒性，控制基因表达的时间的能力就显得尤为重要。

载体 DNA 能通过转化导入宿主生物体中

带有插入 DNA 的载体经转化导入宿主细胞中才能进行扩增。如我们在第 2 章所讨论的，转化（transformation）就是宿主生物体从环境中摄取 DNA 的过程。在一些细菌中，这个过程能自然发生，我们把它们叫做是有遗传感受态（genetic competence）的。*E. coli* 中转化不能自然发生，不过用钙离子处理后，*E. coli* 也可被赋予摄取 DNA 的能力。细菌摄取 DNA 的确切机制尚未知晓，可能是因为钙离子能遮蔽 DNA 上的负电荷，从而使 DNA 能穿过细胞膜。因此据说经钙离子处理的细胞可用于转化。摄取质粒 DNA 的细胞被称作转化子，转化子可以用抗生素来进行筛选，由于质粒对某一种抗生素有抗性，携带质粒的转化子也因而获得了这种抗生素抗性，能在这种抗生素中生长，但不含质粒的细胞则不能生长。

通常转化相对效率并不高，只有一小部分用 DNA 处理的细胞能摄取质粒。正因为转化的效率低，我们有必要对转化子用抗生素进行筛选。用 DNA 处理后，细胞被转移到含有相关抗生素的培养基中，只有摄取了质粒并能稳定携带质粒的细胞才能生长。转化的低效率能保证在绝大多数情况下，每个细胞只摄取一分子 DNA。这个特性使每个转化细胞及其后代只携带单个 DNA 分子，从而能有效地将这个 DNA 分子从转化混合物的其他 DNA 中纯化出来并进行扩增。

用克隆法制备 DNA 分子文库

如果起始 DNA 简单，那么制备它的特异性克隆是轻而易举的事。如果起始 DNA 小（例如，来源于基因组大小仅为 10kb 左右的小病毒），我们只要在限制酶酶切和凝胶电泳后分离 DNA 片段就可以了。在插入到载体之前，不同大小的 DNA 可以从凝胶中回收并纯化。

如果起始 DNA 比较复杂（如人类基因组），要做克隆就不太容易了。在这种情况下，用一种限制酶处理后进行简单的电泳分离，在切割位点之间的平均距离附近有大量不同大小的片段。因此，在这种情况下，我们不如将所有的片段群进行克隆，再分离单个克隆。

DNA 文库（library）就是一群含有不同的 DNA 插入片段的相同载体（图 21-8）。为了构建 DNA 文库，我们选择一种能产生适宜大小的插入片段的限制酶对靶 DNA（如人类基因组 DNA）进行消化。插入片段的大小可以在小于 100 碱基对至大于 1Mb 之间（极大的插入片段通常由 DNA 不完全酶切造成）。将酶切所得的 DNA 与经同一种限制酶酶切的相应载体混合，用 DNA 连接酶连接，这样就能产生大量携带不同插入片段的载体。

不同来源的插入 DNA 能产生不同种类的文库。最简单的文库是基因组文库（genomic library），它采用全基因组 DNA 经单一限制酶切割所产生的片段制备而成。这种文库在制备基因组测序所用的 DNA 中最为有用；另一方面，如果要克隆编码某一特定基因的 DNA 片段，只有在所研究的生物体中非编码 DNA 相对较少时，才适合采用基因组文库。对于基因组较为复杂的生物体来说，由于许多 DNA 插入片段内没有编码序列，这种文库就不太适合了。

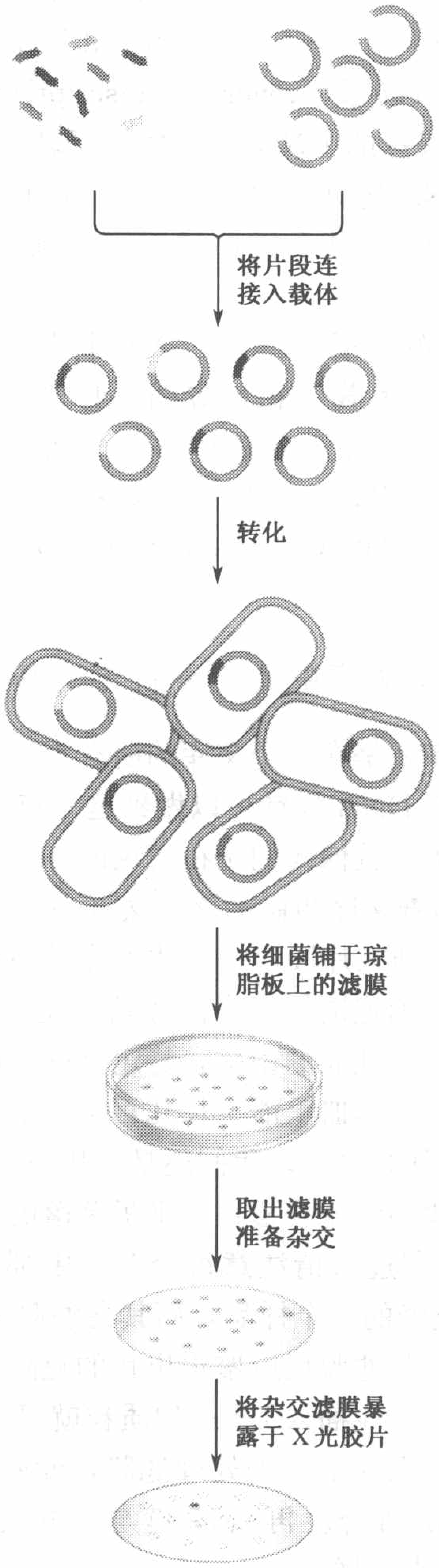


图 21-8 DNA 文库的构建。为了构建文库，我们将同一种限制酶消化所得的基因组 DNA 和载体 DNA 以及连接酶一起进行孵化。然后将最终的杂合载体（每个携带基因组 DNA 的不同片段的细胞，用一种不同的颜色表示）文库导入*E. coli*，接着将细胞铺于琼脂培养基上的滤膜上。菌落长成后，从平板上移走滤膜，进行必要的处理以用于杂交：裂解细胞，使 DNA 变性，再用标记探针孵化滤膜。目的克隆用放射自显影鉴定。

为了在文库中富集编码序列，人们通常使用 cDNA 文库。其制备如下（图 21-9）。在这里，起始 DNA 不再是基因组 DNA，而是由 mRNA 转换来的 DNA。这个由 mRNA 转换成 DNA 的过程叫做反转录（reverse transcription），它由一种以 RNA 为模板来合成 DNA（第 11 章）的特殊的 DNA 聚合酶（反转录酶）催化。用反转录酶处理后，mRNA 能转变成双链 DNA 拷贝，即 cDNA（copy DNA）。此后，构建文库的策略与构建基因组文库一样。cDNA 产物和载体经相同限制酶酶切而使所获得片段连入载体中。

为了从一个文库中分离单个插入片段，人们将整个文库转化进 *E. coli* 细胞中。通常每个转化细胞只含有一个由载体和相应插入片段组成的重组分子。因此，传代后的每个转化细胞就含有文库中某一克隆的多个拷贝。携带任一已克隆的目的序列的细胞所生成的菌落可以被鉴定，目的序列可以被分离。很多方法可以用来鉴定克隆，如下面我们要讲述的是用 DNA 或 RNA 探针杂交来鉴定含有特定插入 DNA 的细胞。

杂交法可以用来鉴定 DNA 文库中的某一特定克隆

克隆基因的一个常用步骤就是在一个文库的所有克隆中鉴定目的基因片段。通过使用和目的基因部分序列互补的 DNA 探针可以做到这一点。这样一个探针可以用来鉴定含有目的基因区段的细胞菌落，就像我们正在讨论的。

用经标记的 DNA 探针筛查文库叫做菌落杂交（colony hybridization）。一个典型的 cDNA 文库含有数千个不同的插入片段，每个插入片段都连接于一个通用载体（见上）。用文库转化合适的菌株后，将细胞铺板于含固体培养基（通常为琼脂，第 22 章）的培养皿上。每个细胞都将长成一个独立的菌落，特定菌落中的每个细胞含有文库中的同一种载体和插入片段（通常每个培养皿上能生长几百个菌落）。

为了便于检测到数量有限的 DNA，我们还是采用 Southern、Northern 印迹中所用的带正电荷的滤膜。这里，我们将滤膜覆盖于平板菌落的上方，滤膜上就留下了每个菌落的细胞（包括一些 DNA）印迹（请注意每个菌落中都有一些细胞留在平板上）。因此，滤膜上包含每个 DNA 克隆的一个样品，而其在滤膜上的位置与其菌落在平板上的位置相对应。这保证当用探针和滤膜反应鉴定出目的克隆后，就很容易通过其对应的菌落位置鉴定出那个菌落，含有相应插入 DNA 的质粒就可以被纯化。

用探针筛选滤膜的反应步骤如下。先处理滤膜，使膜上的细胞裂解，其中的 DNA 被释放出来并在原位与滤膜结合。再将滤膜与经标记的探针孵化，反应条件和 Southern、Northern 印迹实验中一样。

如前所述及第 22 章将讨论的，人们对噬菌体（尤其是 λ 噬菌体）进行了一些改造以用作载体，用噬菌体作载体制备的文库的筛选方法和质粒文库基本相同。不同的只是我们要筛选的是噬菌体在菌苔上生长所形成的噬菌斑，而不是菌落（第 22 章）。

化学合成的寡聚核苷酸

我们在这一章中所描述的好几项技术都需要用到寡聚核苷酸（oligonucleotide），即

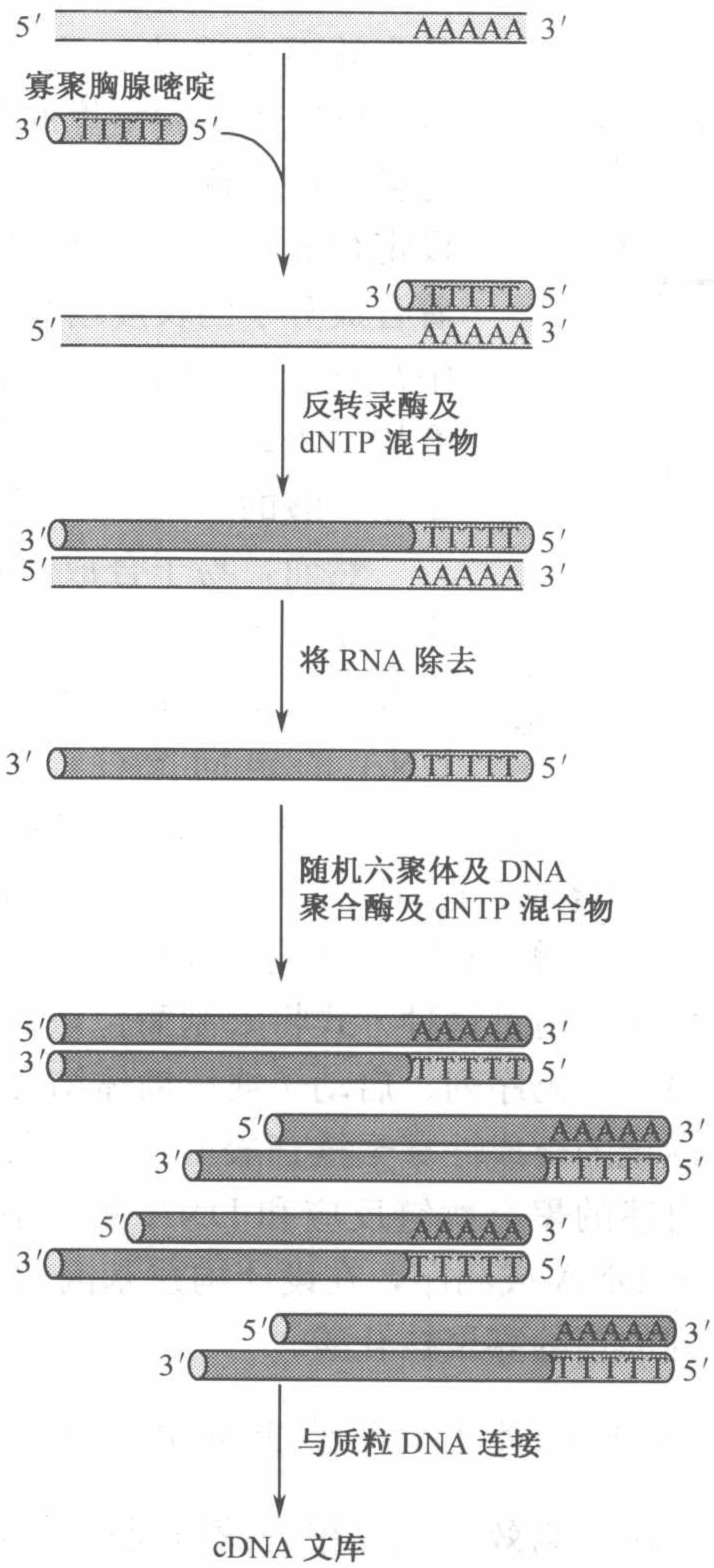


图 21-9　cDNA 文库的构建。RNA 依赖的 DNA 聚合酶——反转录酶将 RNA 反转录成 DNA（cDNA）。第一步（第一条链合成）中，poly T 和 mRNA 的 poly A 尾杂交，起到了引物的作用。反转录酶延伸 dT 引物，产生 mRNA 模板的完整拷贝，形成一条由 mRNA 和它的互补链组成的双链。RNA 用碱（NaOH）处理除去，留下的单链 DNA 就可以作为第二步（第二条链合成）的模板。用随机六核苷酸序列作引物，和 DNA 模板链的不同序列杂交。然后用 DNA 聚合酶延伸引物而生成双链 DNA，这样就可以克隆进入质粒载体（图 21-8）以供制备 cDNA 文库之用。

按照需要合成的短 DNA 片段。虽然 DNA 聚合酶是合成 DNA 分子的最有效机器，DNA 也能通过化学法合成。最常见的化学合成是在 DNA 自动合成仪的固相支持物上进行的。用作添加核苷酸的前体物是一种化学保护了的分子，叫做亚磷酸酰胺（phosphoamidine）(图 21-10)。与DNA 聚合酶催化的链延伸方向相反，化学法合成中 DNA 链的生长是通过在 5′端添加寡聚核苷酸进行的。

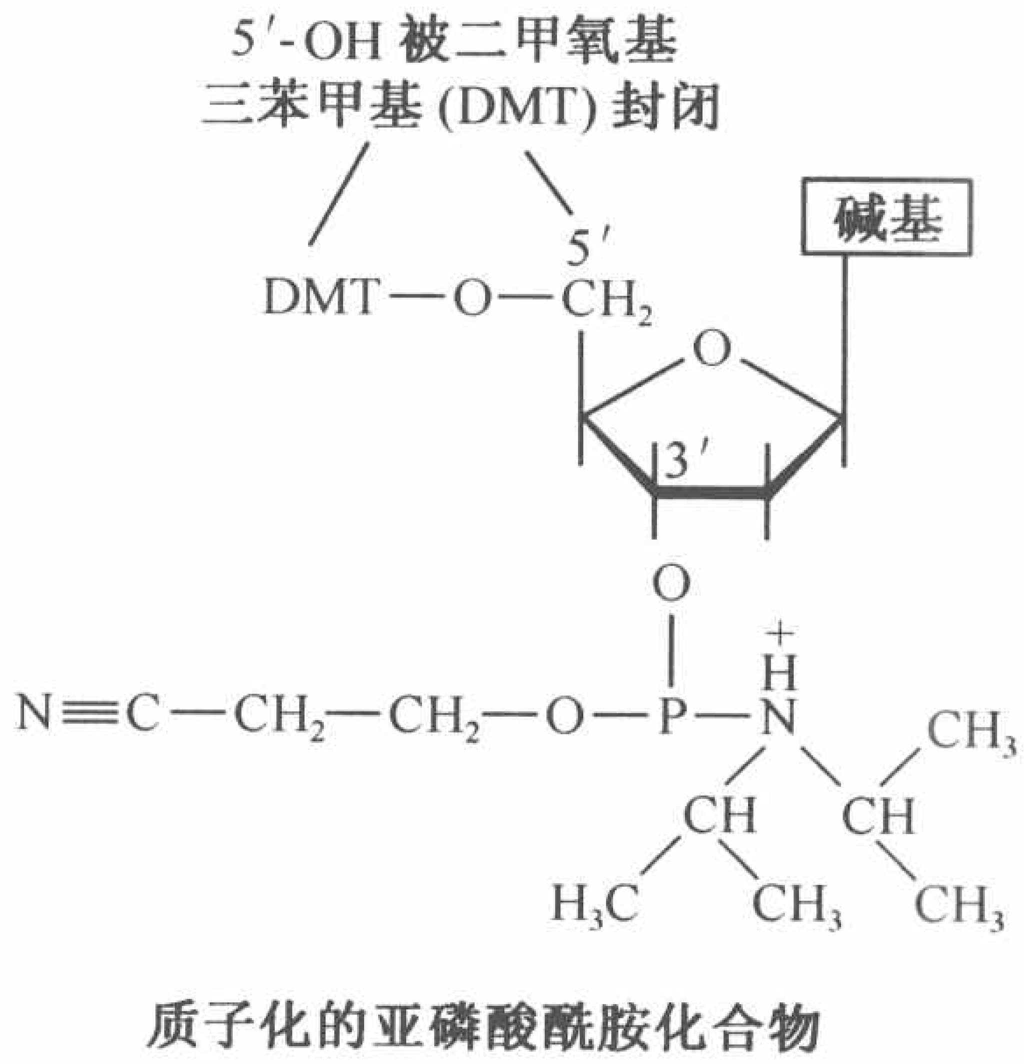


图 21-10 质子化的亚磷酸酰胺化合物。如图所示，5′-OH 被所添加的二甲氧基三苯甲基封闭。

用化学法合成 60～100 个碱基的 DNA 分子是高效而准确的，只需要几个小时。现在 DNA 合成已成了实验室里的常规步骤：研究者们只需将碱基序列输入到操纵 DNA 合成仪的计算机中，设定合成程序，就可以得到任何想要的序列。随着合成分子的长度增长，由于合成过程任一循环中都可能发生内在错误，所得到的终产物就不那么均一了。因此，要大量、准确地合成大于 100 个核苷酸的 DNA 以用作分子分析是比较困难的。

然而，易于合成的短 DNA 分子仍然适于很多用途。例如，含有克隆 DNA 某一区段的某一错配位点的特制寡聚核苷酸可用来产生该克隆 DNA 的定向突变。这种方法叫做定点诱变（site-directed mutagenesis），其步骤如下。寡聚核苷酸与克隆片段杂交后，以克隆 DNA 为模板指导新的 DNA 的合成。这样就产生了含有一个错配位点的双链分子。然后将两条链分离开来，进一步扩增含有所需错配位点的那条链。

人工设计的寡聚核苷酸也能通过这种方式将限制酶位点引入克隆 DNA 中，这可以用来制备一个编码序列与另一编码序列、启动子或核糖体结合位点的融合体。再举一个例子，荧光标记或放射性标记的合成性寡聚核苷酸可以用作杂交的探针。而且，人工设计的寡聚核苷酸在我们要讲述的聚合酶链反应和 DNA 测序中都起着十分关键的作用。因此，为了检测、扩增特定 DNA 或测序，在设计构建基因的新的分子克隆时，都要设计并合成一条含有所需序列的短的寡聚核苷酸。

聚合酶链反应（PCR）通过体外 DNA 的重复复制而扩增 DNA

一种扩增特定 DNA 区段的高效方法是聚合酶链反应（polymerase chain reaction，PCR），它与在宿主细胞中的克隆和繁殖不同，它完全是体外的生化反应。PCR 采用 DNA 聚合酶，以单链 DNA 为模板，以脱氧核苷酸为底物合成 DNA。如第 8 章中提到的，当人工设计的寡聚核苷酸与更长的 DNA 模板发生退火时，DNA 聚合酶将核苷酸加到人工设计的寡聚核苷酸的 3′端。因此，如果一条合成的寡聚核苷酸与和它互补的单链模板退火，DNA 聚合酶就能以它为引物，从 5′→3′延伸而产生双链 DNA。

怎样利用 DNA 聚合酶和 PCR 来扩增特定的 DNA 序列呢？先合成两条单链寡聚核苷酸。一条寡聚核苷酸与要扩增的 DNA 的一条链的 5′端互补，另一条则与另一条链的 5′端互补（图 21-11）。然后使要扩增的 DNA 变性，寡聚核苷酸和它们的靶序列退火。这时，在反应中加入 DNA 聚合酶和脱氧核糖核苷酸底物，DNA 聚合酶使两条引物延伸。这个反应能在 DNA 两条链的目的区域处产生双链 DNA。因此经过 PCR 反应的第一个循环，产生了 DNA 起始片段的两个双链拷贝。

接着，将 DNA 进行另一轮变性，用相同的引物进行 DNA 合成（注意只有两条引物中间的序列是被准确扩增的。）。这样就生成了 4 个拷贝的目的片段。如此重复进行变

性和引物指导的 DNA 合成能使两个引物之间的区域呈几何级数（2、4、8、16、32、64…）扩增。因此原来含量稀少的一个 DNA 片段经过扩增就能够产生大量的双链 DNA（图 21-11）。

从某种意义上说，DNA 克隆和聚合酶链反应（PCR）依据的是同一思路，即 DNA 的重复复制能大量扩增微量 DNA，不管是通过细胞分裂的反复进行还是体外 DNA 合成的循环进行。然而，在克隆中，我们用选择性试剂或其他设置来确定一个现存文库中的被扩增序列；而在 PCR 中，选择性试剂就是这对寡聚核苷酸，它们从一开始就将扩增限定于特定的目的 DNA 序列（框 21-1）。

框 21-1　法医学与聚合酶链反应

设想你在一个法医学实验室里，手里有一个嫌疑犯的 DNA 样本。你想确定涉嫌者的 DNA 是否含有犯罪现场所发现的 DNA 多态性。多态性就是生物群体中，在染色体的共同的、同源的区域，如一个基因中，存在的不同的 DNA 序列（等位基因）。一个多态性可以简单到只是不同生物体染色体的某一位点的单个碱基差异，或者是简单重复序列如 CA（第 9 章）的长度差异。我们想做的是扩增多态性位点周围（包括位点本身）的 DNA，这样我们可以对得到的 DNA 进行测序（见下），测定它是否含有犯罪现场样品中所发现的序列。扩增所得的 DNA 的核苷酸序列（连同对另外的多态性的检查）能帮助确定两个 DNA 样品是否吻合。与传统指纹识别技术鉴定方法相类比，这种鉴定 DNA 序列的鉴定方法被称作 DNA 素描或 DNA 指纹识别。DNA 素描自 1985 年首次运用（美国联邦调查局自 1988 年开始使用这一技术）以来，已被广泛运用于对犯罪现场证据的分析，用来证明犯罪嫌疑人有罪或者是清白。

4 组部分重叠的 DNA 片段揭示核苷酸序列

接下来我们将讨论如何测定核苷酸序列。从某种意义上说，核苷酸测序是选择性研究基因组的最高手段。整个基因组的完整核苷酸序列的测定，例如，已经完成测序的复杂程度不同的一些生物体（从细菌到人类）的基因组序列，使我们能通过使用计算机和相应的运算法则快速、准确地查找任一特定序列。换句话说，在处理核苷酸序列时我们所用的“选择性试剂”是一串我们输入到电脑中的碱基。随着大量可用的基因组序列的增加，利用计算机在同一种生物体中或不同生物体间高精度地搜寻相关序列的拷贝已成为可能。显然，核苷酸测序产生了功能强大的数据库，这一点我们将在后面详述。

DNA 测序的基本原理是按大小分离几组部分重叠的 DNA 分子。这些 DNA 分子 5′端相同，但 3′端有几种不同的可能性。每组 DNA 分子的 3′端是一种特定的碱基。第一组 DNA 分子都以 G 结束，第二组是 C，第三组是 A，最后一组是 T。一组特定的 DNA 分子，以 G 组为例，它们的长度因其 3′端的 G 在序列中的位置不同而不同。因此，这组 DNA 片段能告诉你产生它们的 DNA 分子中何处有一个 G。我们将在后面讲述如何

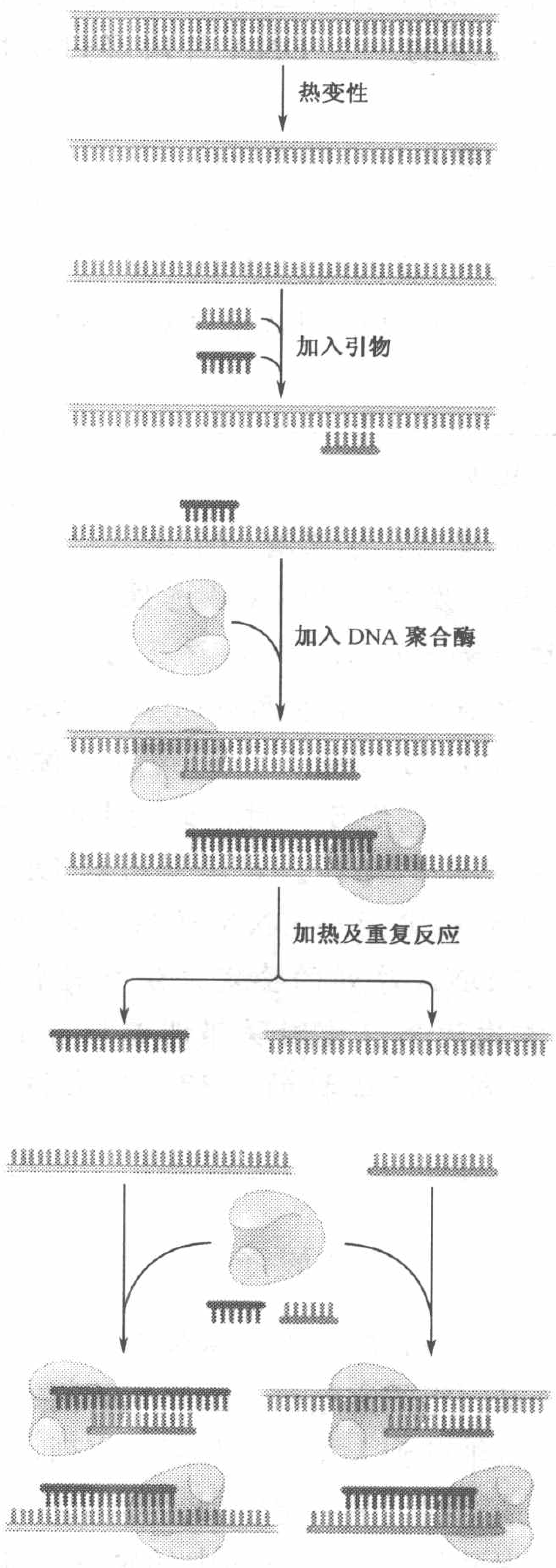


图 21-11 聚合酶链反应。在 PCR 的第一步，将 DNA 模板加热变性，用对应于要扩增的 DNA 序列的边界的寡聚核苷酸引物（深橘红色和暗绿色）与 DNA 模板退火。用 DNA 聚合酶（淡橘红色）在引物处延伸、拷贝单链模板（淡绿色）。然后，再次使 DNA 变性，和引物退火并用作新一轮 DNA 合成的模板。请注意在第二个循环中，引物既能引发依据原始模板的 DNA 合成，也能引发依据新合成的 DNA 的合成。当 DNA 聚合酶延伸与来自于上一轮 DNA（或来自于绿色标记模板的橘红色引物）合成产生的新模板（橘红色标记）退火的绿色标记的引物时，它沿着模板链前进直至其末端（在下图中聚合酶还没到模板的末端）。于是，欲扩增的 DNA 序列在第二轮循环中就被准确合成了。此后，进一步循环变性、引物退火和 DNA 合成（未标出）就能产生处于两个引物之间的 DNA 片段。反应每进行一轮，DNA 的量就呈几何级数增长一次。

产生这些片段（图 21-14）。

这些长度不同的片段可以用聚丙烯酰胺凝胶电泳进行测定。将 G 组片段跑胶能产生一系列条带，每个条带都对应于一个片段，其长度反映了某一个 G 在 DNA 序列中的位置。可以将 4 组部分重叠的 DNA 平行跑胶，产生 4 组能显示 DNA 序列中何处有 G、C、A 和 T 的条带。比较这 4 组条带的位置，就能显示待测 DNA 分子的完整序列。也可以用不同的荧光分别标记 4 组部分重叠的 DNA，将它们作为一个混合物进行电泳，再用荧光检测仪区分。

如何产生这 4 组部分重叠的 DNA 分子呢？有两种方法：一种方法是用放射性标记 DNA 分子的 5′端，再用 4 种不同的化学步骤处理，使它们分别优先在 G、C、A 或 T 处断裂。这种化学方法已不再广泛使用，因此我们也不再赘述。另一种方法采用链终止性核苷酸（chain-terminating nucleotide）和体外 DNA 合成，这种方法至今仍广为使用，也是现代自动测序仪 Sequenator 所依据的技术。

链终止法中，寡聚核苷酸引物定位 DNA 合成的起始位置后，DNA 聚合酶从 DNA 模板中复制出若干拷贝分子。如我们在第 8 章中所见，DNA 聚合酶以 2′-脱氧核苷三磷酸为底物，从 5′→3′合成 DNA。（链终止法依据的是我们第 8 章讨论的 DNA 的酶合成）。链终止法采用特殊的、经修饰的底物，即2′,3′-双脱氧核苷酸（ddNTP），它们的糖环上没有 2′和 3′-OH（图 21-12）。DNA 聚合酶能在延伸中的多聚核苷酸链的 3′端掺入一个2′,3′-双脱氧核苷酸分子，但它一旦掺入后，3′-OH 的缺乏阻止了核苷酸进一步的延伸而使链延伸反应终止（图 21-13）。

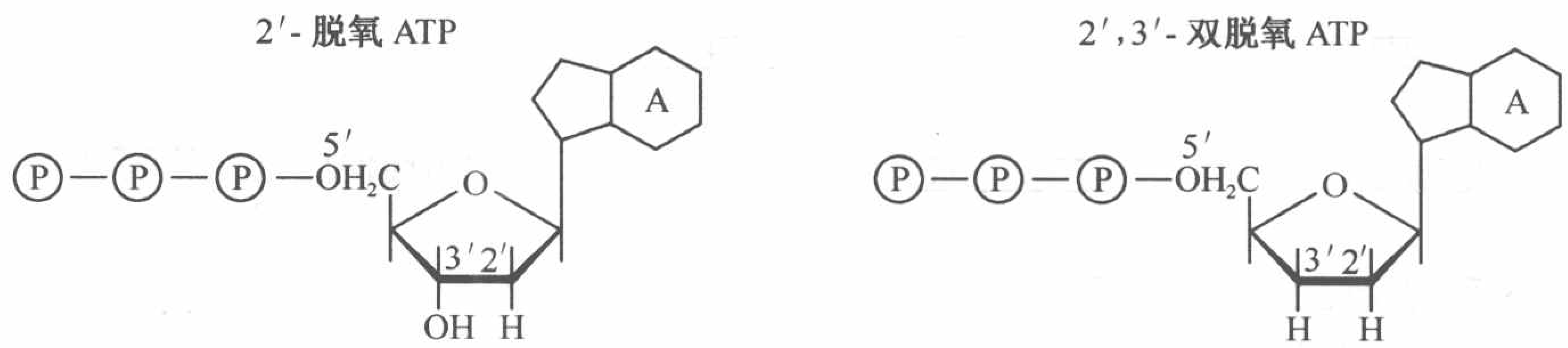


图 21-12 DNA 测序中使用的双脱氧核苷酸。左图是 2′-脱氧 ATP，它能加入到延伸中的 DNA 链中，使另一个核苷酸直接加在它的后面。右图是 2′，3′-双脱氧 ATP，一旦它加入到延伸中的 DNA 链中，它就阻止其他核苷酸加入到同一条链中。

现在设想我们在核苷酸底物混合物中掺入修饰底物2′,3′-双脱氧鸟嘌呤三磷酸（ddGTP），ddGTP 和 2′-脱氧 GTP 的比值为 1∶100。这就使 DNA 在每次 DNA 聚合酶碰到模板链上的 C 时以 1∶100 的频率终止（图 21-14a）。由于所有的 DNA 链从同一点开始延伸，链终止性核苷酸能产生一组部分重叠的多聚核苷酸片段，所有片段 5′端相同但长度不同，因此它们的 3′端不同。因此这些片段的长度指示了模板链中 C 的位置。如果用放射性标记的引物或用荧光添加剂作标签的引物标记这些片段的 5′端，或用荧光标记的 ddGTP 的衍生物标记它们的 3′端，那么用聚丙烯酰胺凝胶电泳时，这组部分重叠的片段就能产生一系列条带，每条条带代表了模板链上的一个 C.（图 21-14b）。如果我们用 ddCTP、ddATP 和 ddTTP 以相似的方式阻止 DNA 的合成，这样我们总共就可以得到 4 组部分重叠的 DNA 片段，它们一起能提供 DNA 的全部核

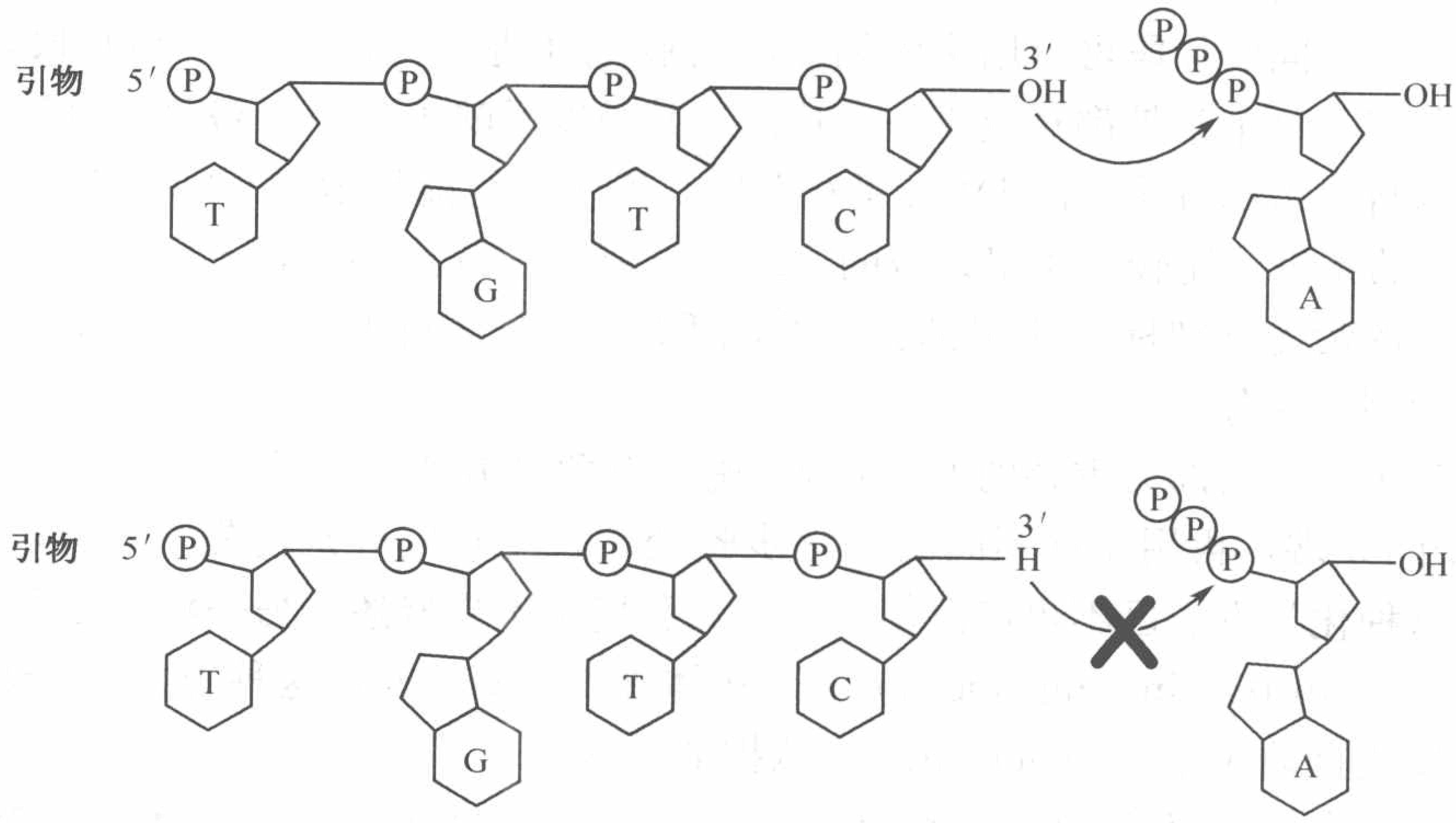


图 21-13　双脱氧核苷酸引发的链终止。上图是一条 3′端正在延伸的 DNA 链，一个腺苷酸正加入到先前掺入的胞嘧啶上。如文中所述，延伸中的 DNA 链的双脱氧胞嘧啶（下图所示）会阻止别的核苷酸的加入。

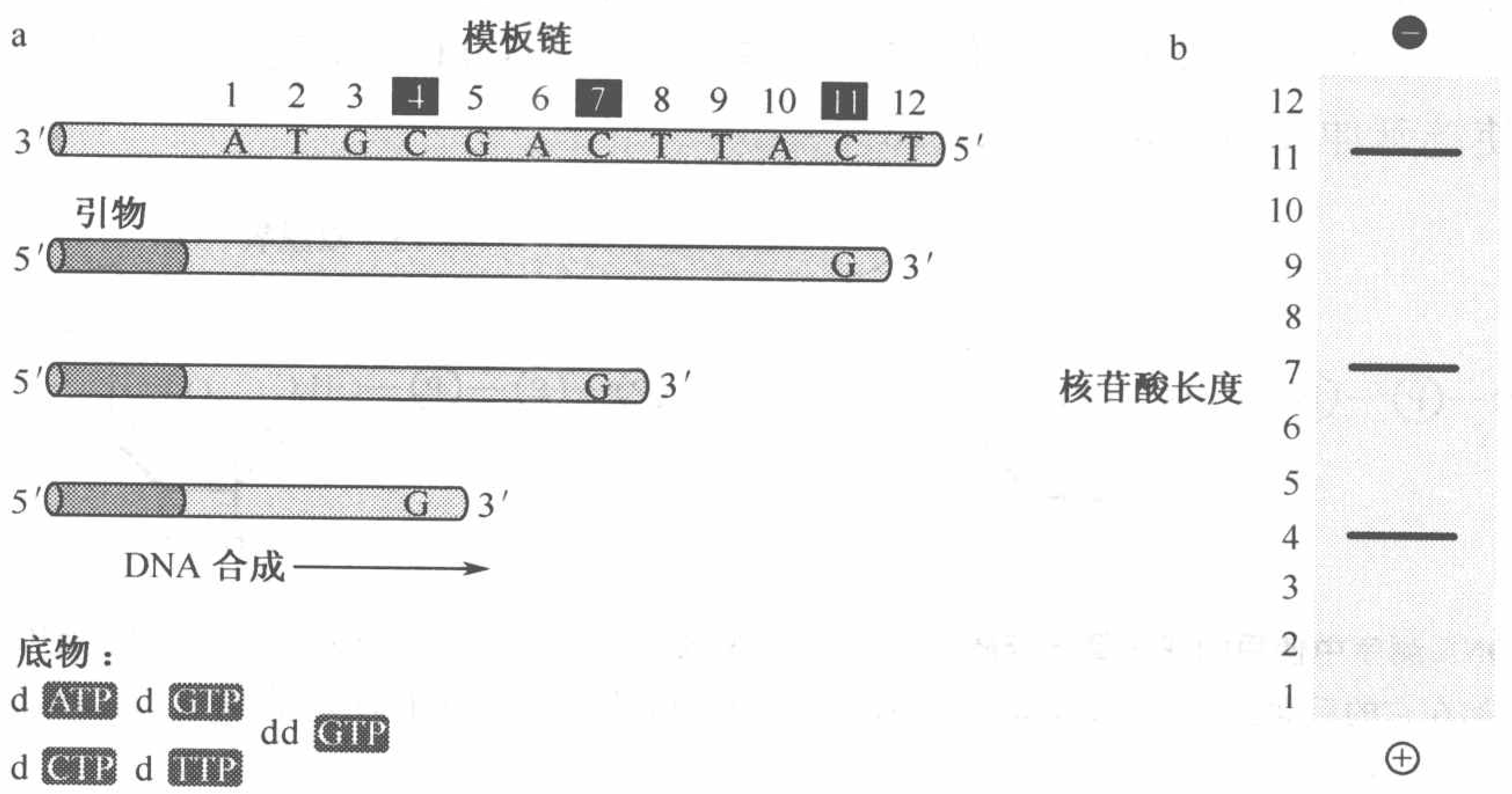


图 21-14　链终止法 DNA 测序。如文中所述，在双脱氧核苷酸存在的情况下，将合成不同长度的 DNA 链。所产生的 DNA 链的长度依赖于 DNA 模板的序列以及反应中掺入了何种双脱氧核苷酸。本图中，a. 最上方是 DNA 模板的序列。在这个反应中，所有的碱基都以脱氧核苷酸的方式存在，但是鸟嘌呤还有双脱氧的形式。因此，当 DNA 链延伸至模板中的胞嘧啶时，一些分子中掺入了 ddGTP，而不是 dGTP。在这种情形下，DNA 链就停止延伸。b. 显示用聚丙烯酰胺凝胶分离得到的片段。凝胶上这些片段的长度反映了进行测序的模板 DNA 中胞嘧啶的位置。

苷酸序列。为了读出这个序列，每种反应所产生的片段都在聚丙烯酰胺凝胶上进行分离（图 21-15）。

下面我们可以看到，这一理论上容易的方法，最开始是为了测定短的、已定义的

DNA 片段的序列而研制的。这一方法经历了一系列技术改进，使全基因组分析成为可能（框 21-2）。

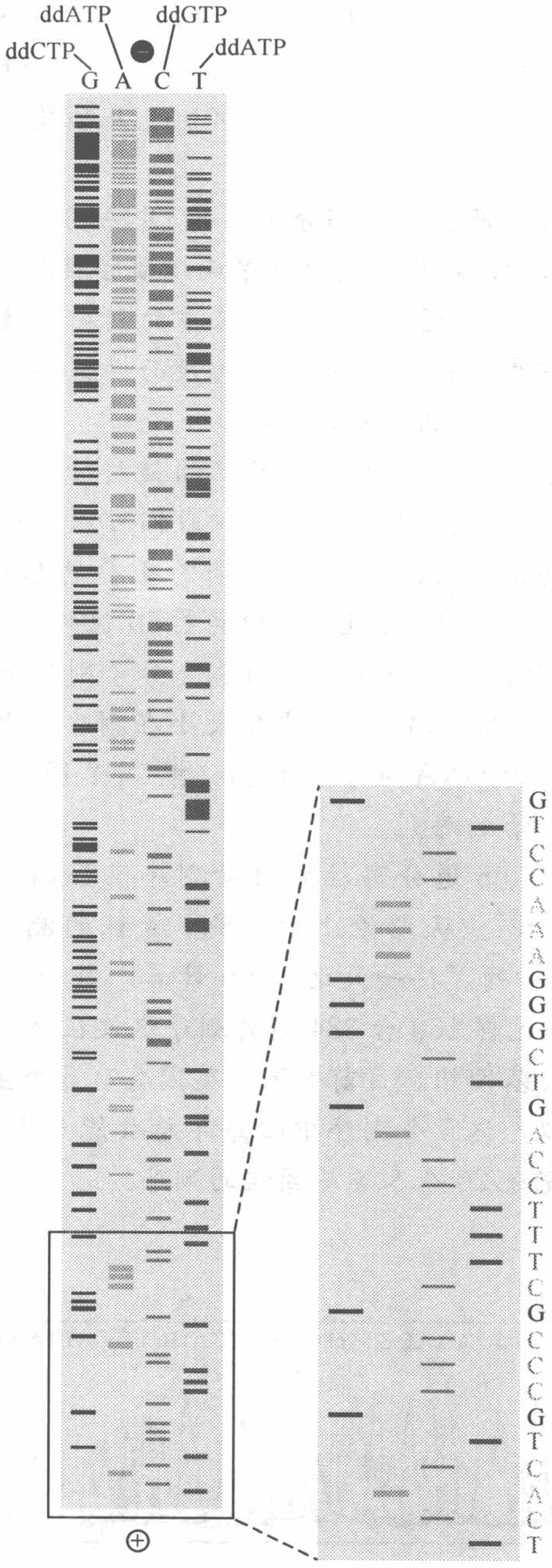


图 21-15 DNA 测序胶。如图所示，DNA 链的长度可以在聚丙烯酰胺凝胶上进行分离测定。DNA 链由每一栏顶部标示的双脱氧核苷酸终止测序反应而产生。从下往上读胶可以得出 5′→3′的序列。

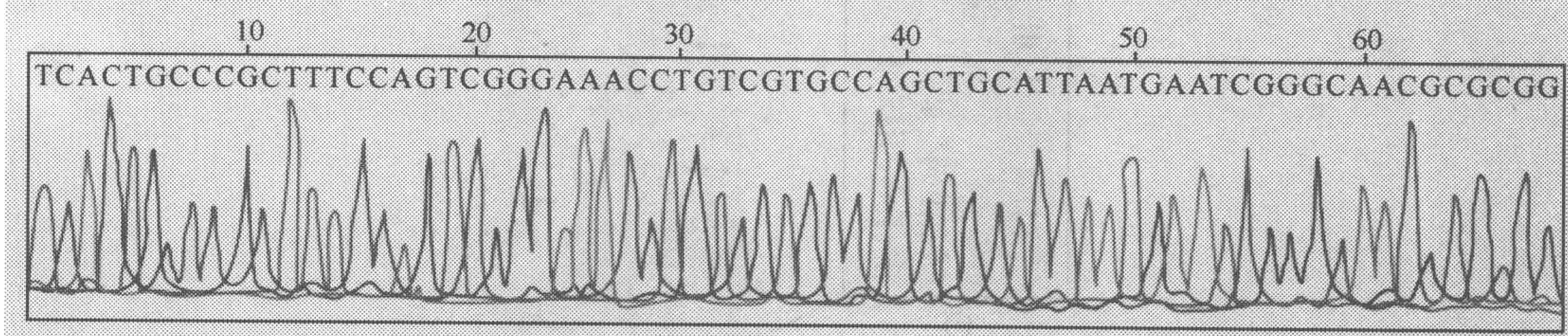
框 21-2　用自动测序仪进行高通量测序

第一次提出人类基因组测序时，看起来还是一件令人生畏、不可企及的事。毕竟整个人类基因组含有 3×10^9 bp，数目惊人，而早期连测定 DNA 短片段的核苷酸序列的方法都相当繁琐。20 世纪 80 年代和 90 年代早期，一个研究者全力以赴，在一两天内也只能产生几百碱基对的序列，也许是 500bp。几项技术上的创新大大提高了测序的速度和可靠性。

如前所述，链终止法能产生 4 组部分重叠的 DNA 分子，这些 DNA 分子之间大小仅相差一个核苷酸。起初，人们采用大的聚丙烯酰胺凝胶电泳分离这些部分重叠的 DNA 分子（图 21-15）。然而，近年来，这些笨重的凝胶已被短分离柱所取代，它们能在 2～3h 内分离部分重叠的 DNA。与繁琐的聚丙烯酰胺凝胶相似，这些可重复使用的短柱子能分离 700～800bp 的 DNA 片段。

DNA 测序的主要技术进展源于荧光标记的链终止核苷酸（fluorescent chain-terminating nucleotide）的使用。从理论上说，我们可以用一种"颜色"标记来自一个片段的每一组部分重叠的 DNA 分子。每组部分重叠的 DNA 分子的颜色对应于它们的最后一个核苷酸。例如，以模板 DNA 中的第 50 位的 T 残基结尾的 DNA 分子可以被标为红色；以第 51 位的 G 残基结尾的 DNA 分子则标为黑色；因此，每个部分重叠的 DNA 分子的大小、颜色都不同。当按大小在测序胶上分离它们时，荧光感受器能检测到每个部分重叠的 DNA 分子的颜色（框 21-2 图 1）。这样，一个柱子在不到 3h 内能产生 600～800bp 的序列。

人们已经研制了含有 96 道分离柱的自动测序仪 Sequenator，最近甚至推出了 384 道分离柱的 Sequenator。从理论上说，384 道柱的测序仪能在数小时内产生 200kb 以上的原始 DNA 序列。按一天运行 9h 计算，每台机器一天能跑 3 次，产生 500kb 以上的序列信息。这样 100 台 384 道的测序仪仅在两个月内就能产生 3×10^9 bp 的序列，相当于一个人类基因组。目前美国、英国共有 5 个主要测序中心，每个测序中心都有多台自动测序仪。这 5 个测序中心每年总共能产生 60×10^9 bp 的原始 DNA 序列。这相当于每年能完成 20 个人类基因组的测序！



框 21-2 图 1　DNA 序列峰图。如正文所述，这一测序反应使用了末端荧光标记的双脱氧核苷酸，反应合成链以毛细管电泳分离，序列中 A、T、C、G 分别以绿色、红色、黑色与蓝色表示。

用鸟枪法测定一个细菌基因组的序列

流感嗜血杆菌（*H. influenzae*）是第一个完成全基因组测序及组装的独立生存的生物体。由于它的基因组小而致密，仅由 1.8Mb 的 DNA 组成，先选择它进行测序是合理的。*H. influenzae* 的基因组被随机剪切成很多随机片段，平均长度为 1kb。这些基因组 DNA 片段接着被重组克隆到质粒 DNA 载体。从单个重组 DNA 菌落中制备 DNA，用前面所讨论的双脱氧法在 Sequenator 上分别测序。这个方法就叫作“鸟枪法”测序。人们挑选随机重组 DNA 菌落，对其处理后进行测序。为了保证基因组中的每个核苷酸都能在最后的基因组组装中出现，人们对 30 000～40 000 个不同的重组克隆进行了测序，产生了总共约 20Mb 的原始基因组序列（通常一个反应产生 600bp 的序列，600bp × 33 000不同的克隆＝共约 20Mb DNA 序列）。这就叫做 10×序列覆盖度(10× sequence coverage)。理论上说，基因组中的每个核苷酸都被测了 10 次。

这个方法可能看起来很繁重，但与最初设想的技术相比，它既快速，花费也少。早期的测序策略要求对细菌染色体物理图上的每一个特定的限制性 DNA 片段进行系统测序。这个步骤的不足之处是大多数已知的限制片段比单个反应所能检测的 DNA 序列要大。因此为了得到基因组任一特定区域的完整序列，人们需要进行更多次的内切酶消化、作图和测序。这些额外的克隆和限制酶作图步骤比随机 DNA 片段的重复自动测序耗时要多得多。换句话说，用计算机组装随机 DNA 序列要比对横跨细菌染色体的整套限制酶酶切片段进行克隆和测序要快得多。

将大约 30 000 条随机基因组 DNA 片段序列的读长输入到计算机中，使用不同的程序对重叠的 DNA 序列进行组装。这个过程的基本思想与填字谜的组装相似。其中已知的单词对未知的相邻重叠单词提供线索。根据匹配序列，随机 DNA 片段可以组装在一起。短的 DNA 序列的组装最终能生成一个连续的组装体，也叫叠连群（见本章的后面部分图 21-17）。

鸟枪法使大基因组序列的部分组装成为可能

从我们前面的讨论中我们知道 600bp 的短 DNA 片段的测序极其快速、高效。实际上，自动测序仪的效率远远超过了我们组装和注释原始 DNA 序列的能力。换句话说，复杂基因组，如人类基因组的全基因组测序的限速步骤是数据的分析，而不是数据的生产本身。随着测序方法变得更加快速有效，这一问题变得愈加严重。现在运用自动机器跑一次即可产生 10 亿碱基对（gigabase pair）的 DNA 序列信息（见下文“向 1000 美元测定人类基因组迈进”部分）。现在我们讨论如何改进测定 *H. influenzae* 的全基因组序列所用的鸟枪法，以便用于更大、更复杂的动物基因组的测序。

人类染色体平均由 150Mb 长的 DNA 组成。因此，一个典型测序反应所产生的 600bp 的 DNA 序列只是一个典型的人类染色体的 0.0004%。所以，为了测定染色体的全部序列，我们有必要得到大量 DNA 短片段的测序结果（图 21-16）。为了实现这一目标，需要从组成人类基因组的 23 条染色体中制备 DNA，并加压通过细针头将它分解成小片段。由单个染色体获得的小片段集群随后被分成池（pool）。通常，对于不同大小

的片段，如 1kb、5kb 或 100kb 长的片段要构建 2～3 个 DNA 池。按照前面所述的方法，这些片段被随机克隆到细菌质粒中而制成文库。

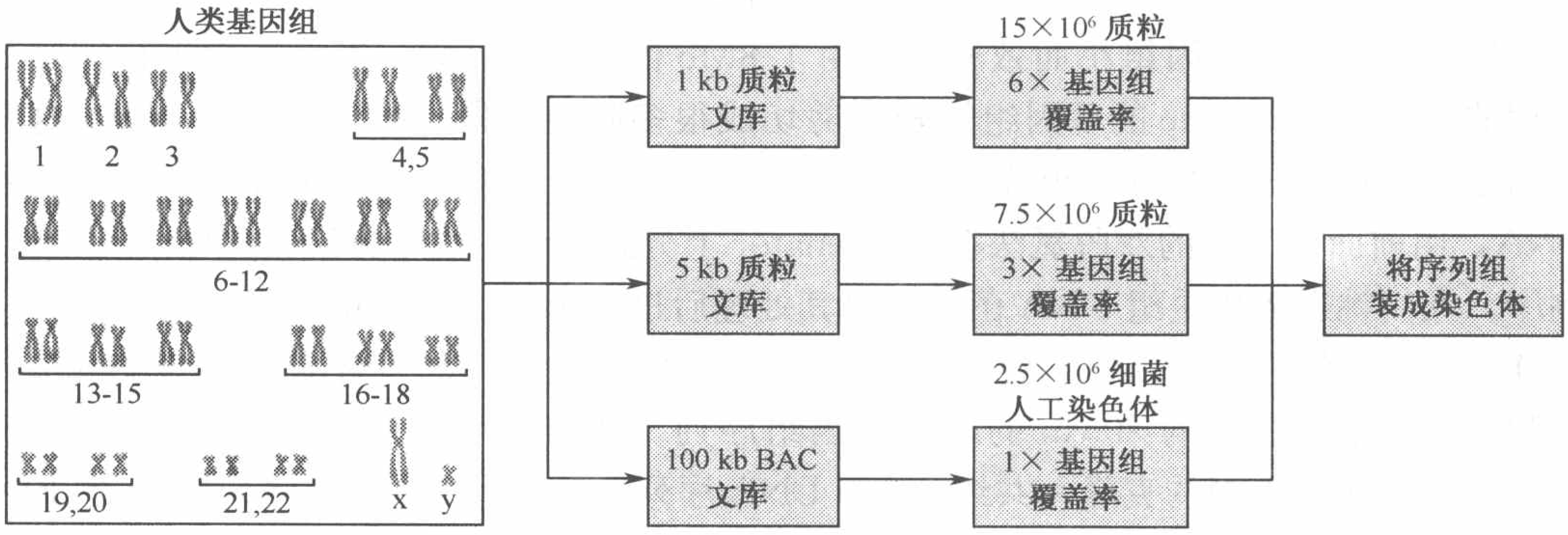


图 21-16 全基因组文库的构建和测序策略。基因组 DNA 短片段的鸟枪法测序后，要构建叠连群。如文中所述，叠连群的延伸可以利用 5kb 和 100kb 等大的插入片段的末端测序来完成。（来源：Hartwell L. et al. 2003. Genetics：From genes to genomes，2nd edition，Fig. 10-13. Copyright © 2003 McGraw-Hill Companies，Inc. Used with permission.）

可以从细菌质粒中快速分离含有人类基因组随机序列的重组 DNA，然后用自动测序仪快速测序。为了保证一条染色体中的每个序列都得到测定，我们平均需要处理两百万个随机 DNA 片段。按每个片段 600bp 长计算，产生的是上十亿个碱基对的序列数据，几乎接近一个典型染色体中 DNA 含量的 10 倍。就像前面细菌染色体测序中所讨论的，如果测序量达到染色体中序列的 10 倍，我们可以认为每个染色体区段都得到了覆盖。

制备“鸟枪法”重组文库和进行大量冗余的随机 DNA 测序看起来很不经济。然而，100 台 384 道自动测序仪几个星期内就能完成一条人类染色体 10 倍覆盖度的测序。这比先分离染色体内的已知区域，再对一组前后重叠的 DNA 片段进行测序的方法快得多。因此，人类基因组测序的技术关键是自动鸟枪法测序（shotgun sequencing）并接着使用计算机组装不同片段。自动测序仪和计算机的联合使用是一套强强联合，最终使人类基因组测序得以提前数年完成。

为了将鸟枪法所得的随机 DNA 短序列组装成叠连群（contig），即大的重叠序列，人们研究开发了先进的计算机程序。含有相同序列的短序列能相互重叠，按重叠序列将片段连接可以形成更大的叠连群（图 21-17）。这些叠连群的大小依所得到的序列的数量而定——所得到的序列越多，叠连群就越大，其间所遗留的空隙就越少。

单个叠连群通常由 50 000～200 000bp 组成，仍然远比典型的人类染色体要短。然而这些叠连群在分析致密基因组中十分有用。例如，果蝇基因组平均每 10kb 有一个基因，因此一个标准的叠连群就包含了好几个连锁基因。不幸的是，更为复杂的基因组中基因密度通常要比这低得多。人类基因组中平均每 100kb 有一个基因，因此一个典型叠连群通常不足以包含一个完整基因，更不用说一系列连锁基因了。接下来我们讨论如何将相对较短的叠连群组装成 1～2Mb 长的支架（scaffold）。

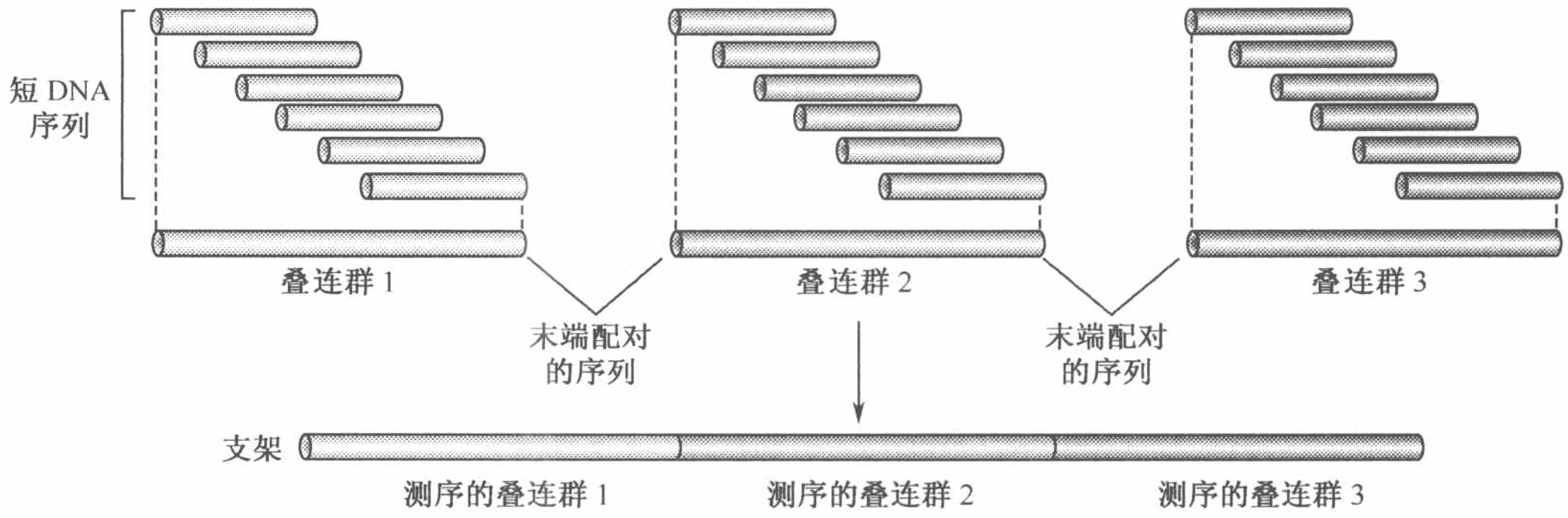


图 21-17 通过测定大的 DNA 片段的末端序列连接叠连群。例如，一个随机的 100kb 的基因组 DNA 片段的末端可能含有和叠连群 1 匹配的序列，而另一个末端含有和叠连群 2 匹配的序列。我们就可以将这两个叠连群放在同一个支架上。（经许可，仿自 Griffiths A. J. F. et al 2002. Modern Genetics，2nd ed.，Fig. 9-29b，© W. H. Freeman.）

末端配对法使大基因组支架的组装成为可能

重复 DNA 的存在是产生更大叠连群的“瓶颈”。一条染色体或一个基因组的随机 DNA 片段中由于含有同一个 DNA 重复序列，尽管互不连锁，看起来也仿佛相互重叠，这使组装过程变得复杂。配对末端测序（paired-end sequencing）可以克服这种困难。这种技术简单易行，十分有效（图 21-18）。

除了制备由 DNA 短片段组成的鸟枪 DNA 文库，还要制备由 3～100kb 长的基因组 DNA 大片段组成的重组文库。以一个来自某一人类染色体的 DNA 样本为例，一部分样本用来制备 1kb 片段，另一部分则用来制备 5kb 片段。最终就构建了两个文库，一个插入片段小，另一个插入片段大（图 21-16）。

制备能在大的 DNA 插入片段的两端与质粒的连接处退火的通用引物。每个随机插入片段末端一次测序能产生 600bp 序列。记录随机插入片段的每个末端序列的来源。一个末端可能与叠连群 A 内的序列匹配，而另一个末端与叠连群 B 内的序列匹配。由于叠连群 A 和 B 与同一个 5kb 片段有共同序列，我们可以推断它们来自于染色体的同一区域。大多数重复 DNA 序列的长度小于 2～3kb，所以来自 5kb 插入片段的“配对末端”序列足以横跨被重复 DNA 打断的叠连群。

前述的方法通常产生小于 500kb 的叠连群。为了得到几个 Mb 以上长序列的数据，有必要从至少为 100kb 以上的 DNA 片段中获取配对末端序列的数据。一种特殊的克隆载体 BAC（细菌人工染色体，bacterial artificial chromosome）可以帮助实现这一点。BAC 中插入的长片段长度可达几百 kb。产生长序列的原理和前述 5kb 插入片段一样。使用引物从 BAC 插入片段的两端获取 600bp 的测序结果，接着寻找能与这些序列匹配的叠连群，由于这些叠连群含有来自同一 BAC 插入片段的序列，我们可以将这些叠连群置于同一支架中。BAC 的使用能使多个叠连群组装在一个含几个 Mb 的支架中（图 21-17）。

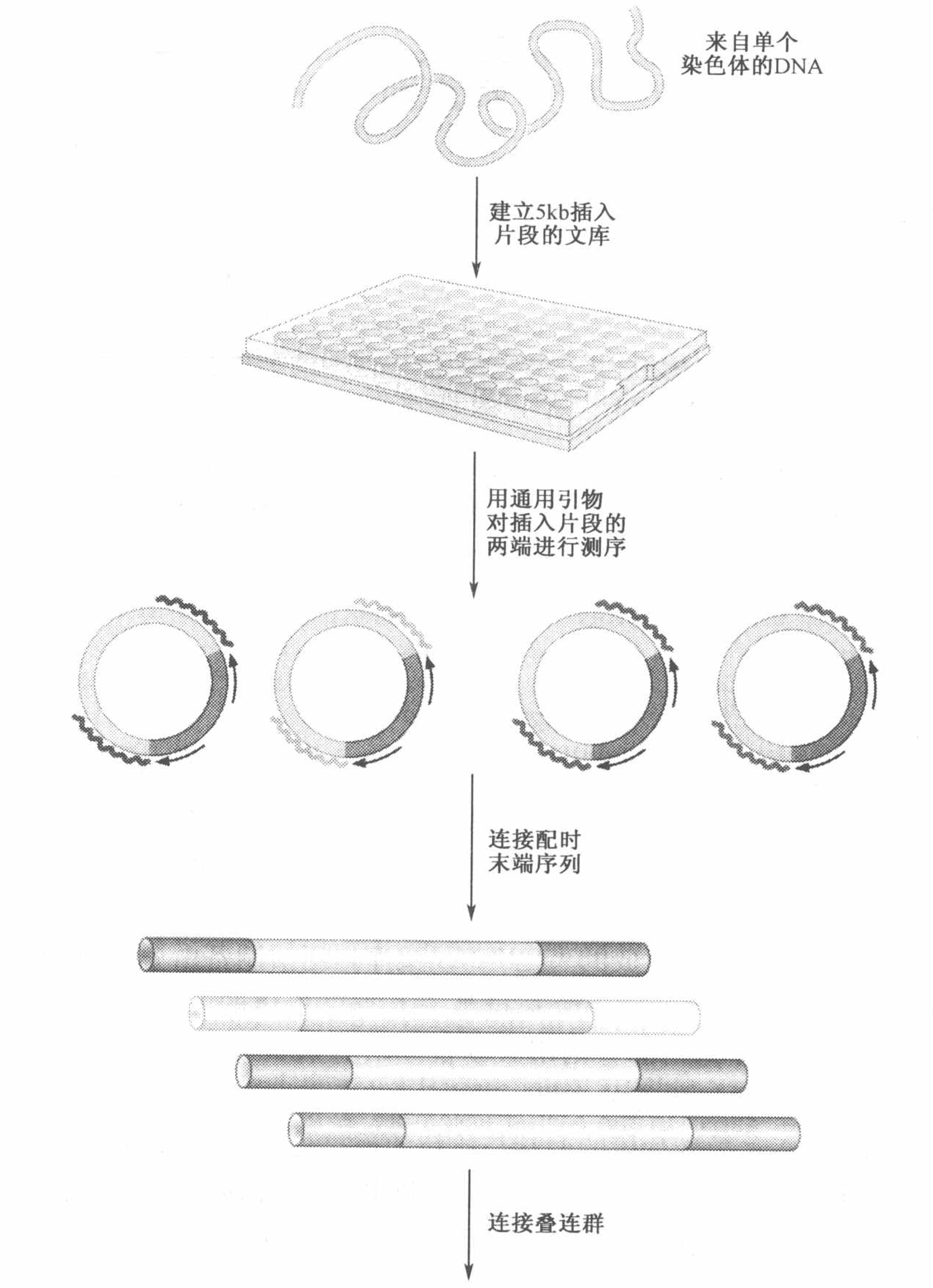


图 21-18 一个包含长度为 5kb 的随机基因组 DNA 插入片段的“鸟枪法测序”文库。平板上每个孔含有一个不同的插入片段。在每个基因组 DNA 片段的两个末端，长度为 600bp 的序列被测定（着色部分）。这些配对末端序列被用来连接不同的叠连群。在本例中，一个 5kb 长度的蓝色基因组 DNA 片段与叠连群 A 和叠连群 B 都匹配。

支架（scaffold）的平均大小可以用来测度基因组组装的质量。若支架的平均大小达到 1Mb 以上，可以认为这个基因组的组装是高质量的。例如，河豚鱼基因组长为 800Mb，其完整的组装序列位于 500 个平均大小为 1.6Mb 的不同的支架上。这一组装已足以用于大多数分析，如所有蛋白质编码基因的鉴定。当 2000 年比尔·克林顿（Bill

Clinton）和托尼·布莱尔（Tony Blair）宣布完成人类基因组测序时，支架的平均大小为 2Mb。就蛋白质编码基因（约 25 000 个基因）而言，这足以用来准确估计人类基因组的基因组成。在过去的七年中，为了解决早期装配时基因组序列上的“gap”，人们采用了各种各样的方法。人类基因组序列图谱基本上已经实际完成，即组成人类基因组的 23 条染色体上都是只由一个序列支架组成。

向 1000 美元测定人类基因组迈进

最早的两个人类基因组的测序工作（一个来自于美国国立卫生研究院，另一个来自于私人企业），耗费超过了 3 亿美元。现在人们开始利用纳米技术进行基因组测序工作，目的是使测序技术速度更快，操作更简便，并且便宜到可以使个人基因组测序用于临床诊断。现在，第一代高通量的纳米技术测序仪已经可以用于测序方面的研究。

454 生命科学测序仪可以在一个 4h 的反应中产生将近 400Mb 的序列信息。它的基本原理很巧妙。首先将短的 DNA 片段（基因组或 cDNA 等）与微珠混合。混合物需要充分稀释，以保证一个 DNA 分子只结合到一个微珠上。然后，把这些结合了 DNA 的微珠分散在一个含有 400 000 个孔的硅板上（每一个孔都达到了皮升级别）。因为尺寸极小，所以每个孔只能容纳一个微珠。然后直接对微珠结合的 DNA 进行 PCR 进行，使每一个 DNA 分子得到扩增。每一个孔扩增产生后就会产生均一的 DNA 分子群，这些 DNA 分子群可以作为模板来进行下一轮的包括生物荧光蛋白质和 DNA 聚合酶在内的 DNA 合成反应（图 21-19）。第二轮 DNA 合成是这样进行的，使 dATP、dGTP、dCTP、dTTP 分别与硅板反应，在每个信号读取峰之间进行洗脱循环。脱氧核糖的加入依赖于模板 DNA 中的互补碱基，并且引起焦磷酸的脱离。焦磷酸的释放促进了酶促反应的发生，产生可被连接在计算上的微处理器检测到的发光脉冲。发光脉冲可以指示出每一轮合成反应后每一个微孔内掺入了什么样的核苷酸，这样就可以获得所有 400 000微孔中所包含的 DNA 序列。核苷酸持续加入直到一个 DNA 片段中大约有 200～250 个碱基被确定。

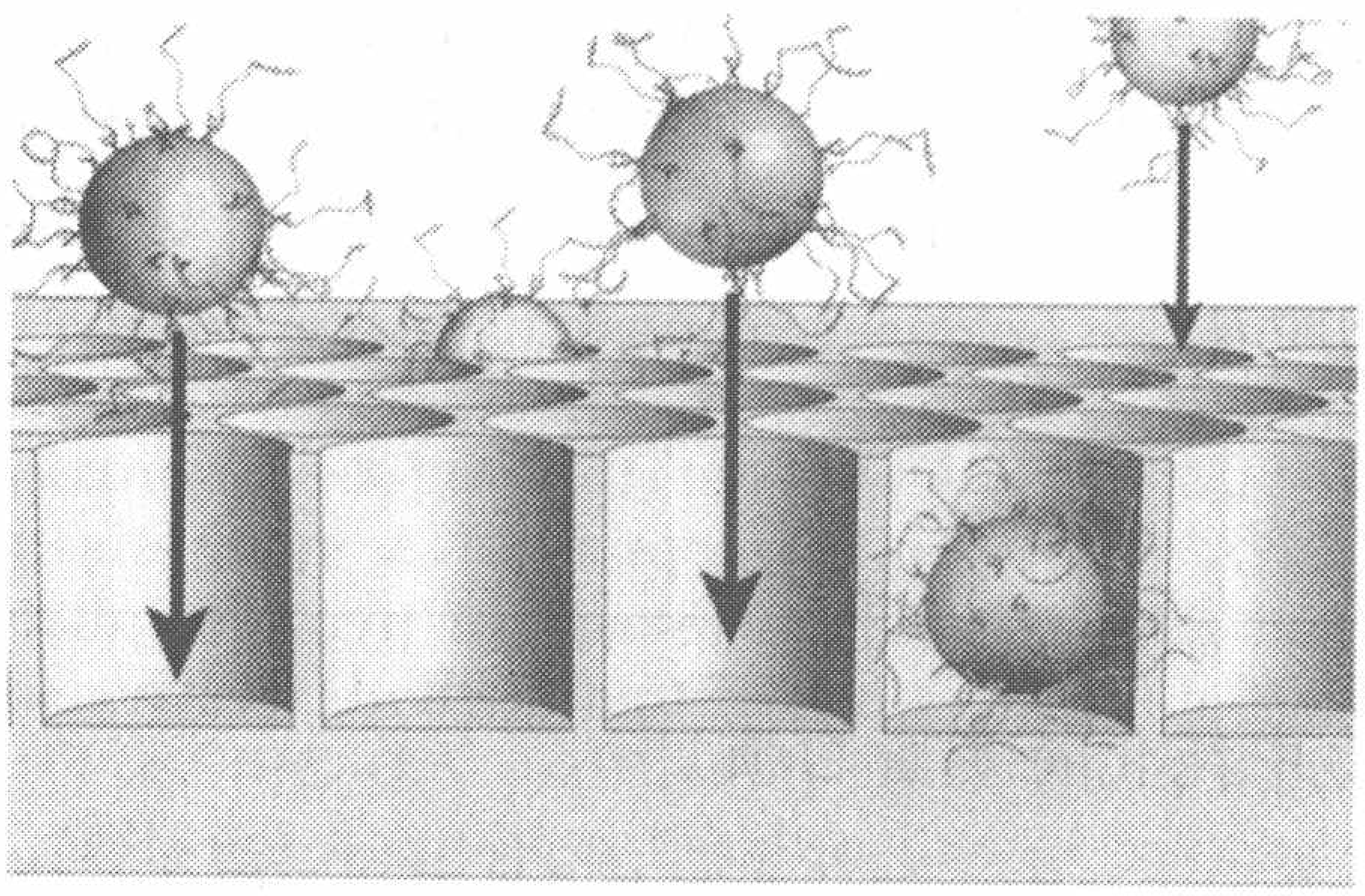


图 21-19 454 测序装置的微孔示意图。每一个孔含有一个微珠，微珠带有扩增的 DNA 序列。连续的多轮测序反应通过焦磷酸的释放和光信号来检测。正文中有进一步的方法描述。（经许可影印自 Margulies M. et al. 2005. *Nature* 437：376-380，Fig. 1a© Macmillan.）

利用 454 测序技术，已经完成了本书第一作者的基因组序列测定（因为某些原因，这家公司似乎对其他作者的基因组没有兴趣）。按照每一轮反应 100Mb 来算，完成沃森的基因组 1 倍覆盖度只需要 30 个反应（一台机器用 2～3 周时间）。根据现在的状态计算（以这篇文章的写作为准），完成一个人的全基因组测序需花费大约 10 万美元，仅仅是第一个人类基因组测序花费的一小部分。为了确保每一个基因都能被覆盖到，就需要提高覆盖的倍数，这需要将费用增加到五十万美元左右。此外，所获得测序数据并不能保证一个新基因组的组装。而美国国立卫生研究院已完成的人类基因组序列此时可用作模板。454 测序仪所读到的所有 200～250bp 的序列都要通过与基因组完成图的比对进行鉴定，直到沃森的基因组中每一个基因的多态性被鉴定出来。所以，测定一个人的基因组的含义已经发生了变化。因为在我们已经有了一个完成了的全基因组序列的基础上，只要获得大量的短序列，就可以获得一个个体的唯一的遗传构成综合图谱。

下一代的测序仪正在接近 1000 美元测定一个基因组序列的目标。Solexa 公司（现在是 Illumina 公司的一部分）开发了一种仪器，该仪器可以在一个反应中获得 4 亿个读长为 30bp 的序列。它的基本原理和前面的 454 测序仪很相似。不同的是，在它的测序反应中单个 DNA 分子是结合在玻片上的 。一定循环次数的 PCR 反应使每个 DNA 分子扩增大约 1000 个拷贝。随后进行 DNA 合成反应，并通过焦磷酸的释放来实现检测。然而，这种方法有两个主要限制，使其测序的长度只有 30～50bp，而不是 200～250bp。第一，DNA 样品是以无序的方式排列在玻璃板上的，所以在合成过程中对所发射信号的检测就比较难；第二，454 测序仪使用的微珠扩增得到的 DNA 分子拷贝数至少是 Solexa 测序仪扩增所得到拷贝数的 10 倍。因此，与 454 测序仪相比，Solexa 测序仪具有较低的信噪比。

尽管有上述这些不利因素，Solexa 测序仪的表现还是相当惊人的：一个测序反应，它可以产生超过 1Gb 的 DNA 序列。因此，只需要三个测序反应，就可以得到 1 倍覆盖度的人类基因组序列，而每个测序反应的花费只有 3000 美元多一点。也就是说，1 倍覆盖度的人类基因组序列花费约 1 万美元。但是，Solexa 的问题在于其读长很短，可能不会产生与参考基因组唯一的匹配。这个问题在一定程度上，可以通过产生更大覆盖度（更多的读序）来克服，更重要的是可以通过提高对随机 DNA 样品的检测来提高读长。从技术上讲，把读长提高到 50bp 是可行的，实现这一点将极大提高测试序列和装配好的标准序列间比较的能力。

蛋白质

可以从细胞抽提物中纯化特定蛋白质

个体蛋白质的纯化对于研究它们的功能来说是至关重要的。尽管有时候也能在复杂混合物中研究一个蛋白质的功能，但这些研究可能导致结果难以解释。例如，如果你在蛋白质的天然混合物（如细胞裂解产物）中研究某一特定 DNA 聚合酶的活性，其他 DNA 聚合酶和辅助蛋白可能部分或全部的造成你所观察到的 DNA 合成活性。由于这

个原因，蛋白质的纯化对于研究它们的功能来说是至关重要的。

每个蛋白质都有其独特的特性，使每个蛋白质的纯化方法各异。这一点与 DNA 的纯化不同。所有的 DNA 有相同的螺旋状结构，只能从它们的精确序列上加以区分。蛋白质的纯化往往利用其独特的特性，包括大小、电荷、形状以及在很多情况下运用功能将其区分。

每一种蛋白质的纯化都需要一种特殊的分析方法

为了纯化一种蛋白质，我们需要针对该蛋白质的独特分析方法。DNA 的纯化几乎总是使用同一种分析，即与它的互补序列杂交。你将在免疫印迹章节中了解到，我们可以用相同的方式用抗体来检测特定蛋白质。在很多情况下，使用一种更直接的检测蛋白质功能的方法更为方便。例如，对于某一特定的 DNA 结合蛋白来说，可以通过检测它与相应 DNA 的相互作用对它进行分析（如在本章末尾“核酸-蛋白相互作用”部分描述的胶迁移检测）。同样，在分析 DNA 或 RNA 聚合酶时，可以在粗提物中添加相应的模板和放射性标记的核苷酸前体物质（标记方法与前述 DNA 的标记类似）。这种类型的分析叫作掺入检测（incorporation assay）。如我们在框 8-1 所讨论的掺入检测在监测催化合成大分子聚合物，如 DNA、RNA 或蛋白质的许多种酶的纯化与功能分析中十分有用。

制备含有活性蛋白质的细胞抽提物

几乎所有蛋白质纯化的起始物质都是细胞抽提物。不像 DNA 能抵抗温度的变化，蛋白质一旦从细胞中释放，轻微的温度波动就能使它们变性。由于这个原因，大多数抽提物的制备和蛋白质的纯化是在 4℃下进行的。细胞抽提物可以用很多不同的方法来制备，我们可以用去污剂、机械剪切力、低渗盐离子处理（使细胞吸收水分而胀破）或压力的迅速改变来裂解细胞。这些方法总的目标都是弱化和裂解细胞膜，使蛋白质得以释放。有时，在应用剪切力（通常使用咖啡研磨器或搅拌器，就像厨房中使用的那种）之前，要先在极低温度下将细胞冷冻。

柱层析法可以将蛋白质分离

最常用的蛋白质纯化方法是柱层析法（column chromatography）。在这种方法中，蛋白质组经充满经过适当修饰的聚丙烯酰胺或琼脂糖凝胶小珠子的玻璃柱而分流。层析柱分离蛋白质有很多方式，每种分离技术根据的是蛋白质的不同特性。这里我们描述 3 种基本方法，前两种方法分别根据蛋白质的电荷或大小对它们进行分离。图 21-20 对这些方法做了概括。

离子交换层析　这种技术根据蛋白质的表面离子电荷的不同，采用经带正电荷或负电荷化学基团修饰的珠子对蛋白质进行分离。在低盐缓冲液中，和珠子相互作用较弱的蛋白质（例如，微带正电荷的蛋白质与带负电荷的柱子的结合）能被洗脱。如果蛋白质和珠子相互作用较强，洗脱时就需要更高的盐离子浓度。在两种情况下，盐离子封闭了带电区域而使蛋白质从珠子上洗脱下来。由于每个蛋白质表面具有不同的电荷量，它

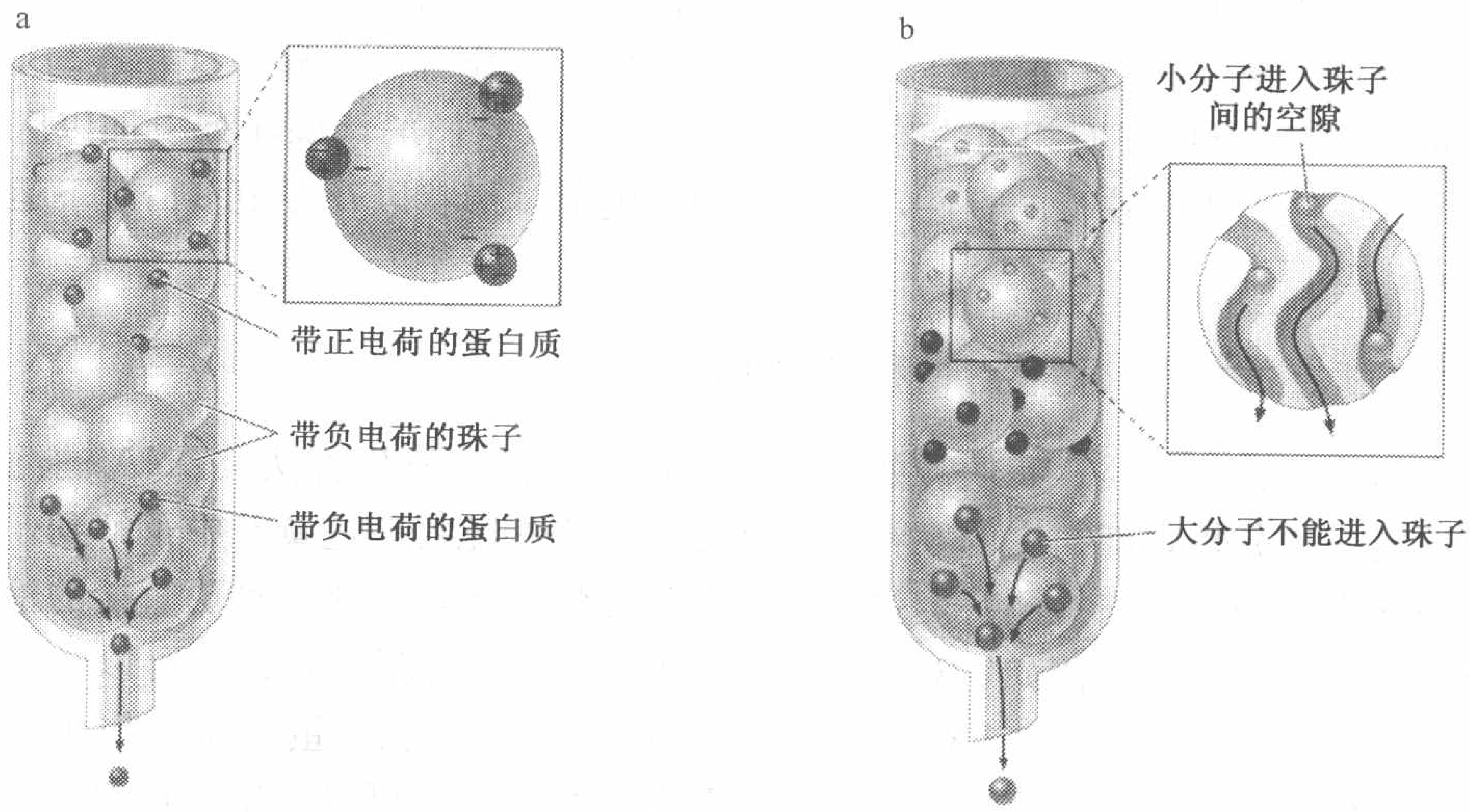


图 21-20 离子交换层析和凝胶过滤层析。如文中所述，这两种常用的层析法分别根据蛋白质的电荷或大小对它们进行分离。在这些方法中，玻璃试管中充满了珠子，蛋白质混合物就流经这一基质。珠子的特性是蛋白质分离的基础。a. 本例中，离子交换层析柱珠子带负电荷，因此带正电荷的蛋白质能与它们结合而滞留于柱子上，而带负电荷的蛋白质流出柱子，提高珠子周围缓冲液的盐浓度，通过盐离子与蛋白质竞争柱上的负电荷而将与柱子结合的蛋白质洗脱；b. 凝胶过滤珠子含有液相空隙，小蛋白质由于能进入这些孔隙而延缓了流经柱子的过程，大蛋白质不能进入珠子而直接流出柱子。

们将在对应的特定盐离子浓度下依次从层析柱上洗脱下来。逐渐增加洗脱缓冲液中的盐离子浓度，可以将带有相似电荷的蛋白质分离成不同组分。

凝胶过滤层析 这种技术根据蛋白质的大小和形状来分离它们。这种层析中所用的珠子没有带电荷基团，但是含有不同大小的孔径（与聚丙烯酰胺或琼脂糖凝胶中使DNA 分子通过的微孔相似）。小蛋白质能进入所有孔径，流经柱子的距离长，洗脱时间也长（换句话说它们需探索更多的空间）；大蛋白质流经柱子的距离短，洗脱速度快。

在不同的盐离子浓度或洗脱时间下收集每种柱子的层析组分，并分析其中的目的蛋白。收集活性最高的组分，然后做进一步的纯化。

将蛋白质混合物通过许多不同的层析柱，它们的纯度将逐渐提高。虽然采用一种层析柱重复分离含有目的蛋白（由该蛋白质的特异分析法测定）的组分很少能完全纯化那种蛋白质，但是一系列的层析步骤可以产生含有许多某种蛋白质分子而极少其他蛋白质分子的组分。例如，尽管有很多蛋白质能被高盐离子溶液从带正电荷的柱子中洗脱下来（这些蛋白质带有高负电荷），或从凝胶过滤柱中缓慢洗脱（这些蛋白质分子相对较小），但同时满足这两种要求的蛋白质就为数不多了。

亲和层析能使蛋白质纯化更加快速

人们通常根据某一蛋白质的特性来更加快速地纯化这种蛋白质。例如，如果你知道某一蛋白质通过和 ATP 结合而起作用，我们就可以使用含有 ATP 结合珠子的层析柱。由于只有与 ATP 结合的蛋白质能与层析柱结合，大多数蛋白质由于不能结合 ATP 而

直接流出柱子。通过加入含浓度逐步提高的 ATP 溶液，这些 ATP 结合蛋白可进一步依据其 ATP 亲合力的大小而被分离（所需 ATP 浓度越高，亲合力越大。）这种纯化方法就叫做亲和层析（affinity chromatography）。层析柱上也可以结合其他试剂来快速纯化各种蛋白质。例如，用结合特定 DNA 序列的层析柱来纯化 DNA 结合蛋白质，或采用能与目的蛋白相互作用的特定蛋白质的层析柱来纯化目的蛋白。因此在开始纯化之前，我们有必要考虑靶蛋白的已知信息并尽量利用这些信息。

免疫亲和层析（immunoaffinity chromatography）是最常见的一种蛋白质亲和层析方法。这一方法中，珠子上结合了一种靶蛋白的特异性抗体。理想的抗体只与靶蛋白相互结合，其他蛋白质都流出柱子。结合于层析柱的蛋白质可以用盐，pH 梯度溶液或在某种情况下用温和去污剂洗脱下来。这一方法的主要困难在于抗体经常与靶蛋白结合如此紧密以致只有将蛋白质变性才能将它洗脱下来。由于蛋白质变性通常是不可逆的，以这种方式得到的蛋白质可能没有活性，因此不是很有用。

为了便于蛋白质的纯化，可以对它们做一些修饰。通常包括在靶蛋白的起始（N 端）或末端（C 端）添加氨基酸短序列，这些添加序列或“标签”可以用分子克隆方法产生。肽段标签能赋予修饰蛋白质一些已知特性，从而有利于它们的纯化。例如，在蛋白质的起始端或末端添加 6 个组氨酸能使它们与带有 Ni^{2+} 离子的层析柱紧密结合，这一特性在一般蛋白质中并不常见。另外，人们鉴定了能与任一蛋白质结合的特定表位（epitope，可被抗体识别的7～10 个氨基酸序列）。使用免疫亲和层析和添加表位的特异性异种抗体可以纯化被修饰蛋白。重要的是，这些抗体和表位的选择要使它们在一种条件下能高亲和性（如 Ca^{2+} 离子存在的情况下）地与层析柱结合，而在另一种条件下（没有 Ca^{2+} 离子）则容易被洗脱。这样就避免了使用能导致蛋白质变性的洗脱条件。

免疫亲和层析还能用来从粗提物中快速沉淀某一特定蛋白质（以及任一与它紧密结合的蛋白质）。在这种情况下，通过将抗体结合于柱层析中使用的珠子而得到沉淀。由于这些珠子比较大，它们迅速地与抗体以及结合于抗体的其他蛋白质沉到试管底部。这个过程叫做免疫沉淀（immunoprecipitation），用于从粗提物中快速纯化蛋白质或蛋白质复合物。虽然免疫沉淀所得到的蛋白质还不完全纯，但这个方法能测定哪些蛋白质或其他分子（如 DNA，见本章后面的染色质免疫沉淀）与靶蛋白结合在一起。

用聚丙烯酰胺凝胶分离蛋白质

蛋白质既没有均一的负电荷，也没有均一的二级结构。它们由 20 种不同的氨基酸组成，有的氨基酸不带电荷，有的带正电荷，还有的带负电荷（图 5-4）。如我们在第 5 章中所讨论的，蛋白质有大量的二级、三级结构，通常以多价复合物的形式存在（四级结构）。然而，如果用强离子去污剂十二烷基硫酸钠（sodium dodecylsulphate，SDS）和还原剂，如巯基乙醇处理蛋白质，蛋白质的二级、三级和四级结构通常能被消除。蛋白质一旦被 SDS 包被就表现为一个无结构的多聚物。SDS 离子包被蛋白质的多肽链，使它携带均一的负电荷。巯基乙醇还原二硫键，破坏半胱氨酸残基间形成的分子内或分子间二硫键。在这些情况下，蛋白质的电泳迁移率取决于其多肽链的长度，电泳就可以像分离 DNA 和 RNA 混合物一样来分离蛋白质混合物（图 21-21）。电泳后，可以利用与蛋白质结合的染色剂，

如考马斯亮蓝（Coomassie brilliant blue）使蛋白质显色。如果不用 SDS 处理，电泳分离蛋白质不是根据蛋白质的分子质量，而是净电荷和等电点等特性（见下文）。

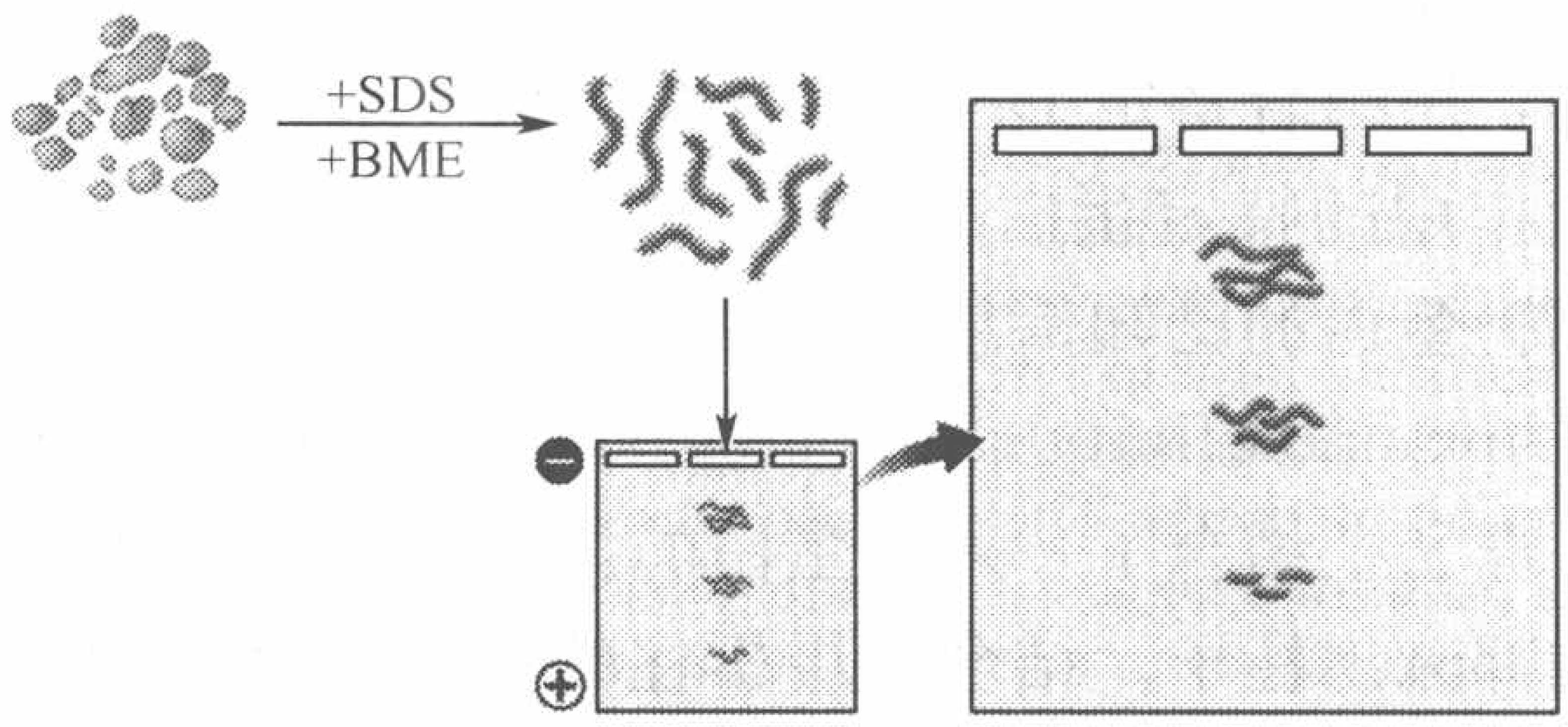


图 21-21 SDS 凝胶电泳。图示表明一个含有三种不同大小蛋白质的混合物（通常需分析的是更复杂的蛋白质混合物），加入 SDS（用红色标记）和 β-巯基乙醇变性后，每一个蛋白质都带有均一的负电荷。从而可以按分子质量大小对蛋白质进行电泳分离。

抗体能显示电泳分离的蛋白质

蛋白质虽然和 DNA、RNA 很不相同，但一种分离蛋白质的方法——免疫印迹（immunoblotting）的原理却和 Southern、Northern 印迹杂交相似，它能在数千种蛋白质中检测并分离某种蛋白质（图 21-22）。事实上，免疫印迹也仿效之前的实验而叫作 Western 印迹。在免疫印迹中，通过电泳分开的蛋白质被转移到非特异性结合的滤膜上。就像 Southern 印迹一样，蛋白质被转移到膜上，蛋白质在膜上的位置与它的凝胶上的位置一样。就像我们讨论过的其他印迹方法一样，一旦蛋白质结合到膜上，所有的非特异性的结合位点需要通过使用与待检测蛋白质不相关的蛋白质溶液进行封闭（通常选择脱脂奶粉来封闭，因为它含有很多白蛋白）。接着把滤膜与特异性识别目标蛋白质的抗体进行孵育，抗体只能和滤膜上的目的蛋白质结合。最后，抗体中人为加入的显色酶（或者加入能够与一抗结合的二抗）使膜结合的抗体显影。Southern、Northern 和

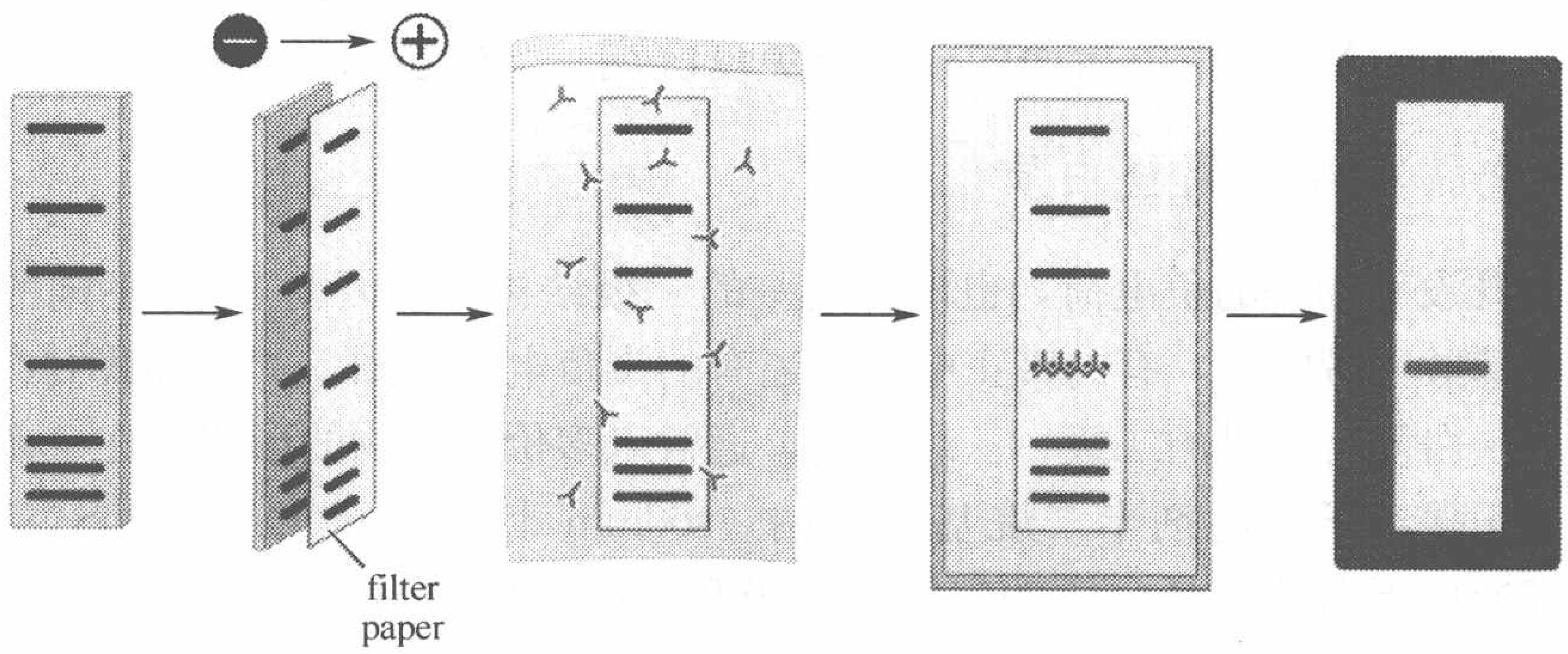


图 21-22 免疫印迹。电泳分离蛋白质以后，把凝胶上的蛋白质相应的转移到滤纸上（同样靠电场作用）。封闭非特异结合位点，在滤纸上加入针对目标蛋白质的抗体。抗体结合的位置通过加入与底物作用产生荧光的酶来检测。

免疫印迹的一个共同特点是利用选择性试剂（selective reagent）使复杂混合物中的特定分子可视化。

蛋白质分子可以被直接测序

蛋白质也可以被测序，即肽链中的氨基酸线性排列序列可以被直接测定，尽管它们的测序要比核酸测序复杂得多。两种常见的蛋白质测序法为 Edman 降解法和随机质谱法，Edman 降解法中需使用自动蛋白质测序仪。蛋白质测序对于鉴定蛋白质是十分有用的。而且，由于我们已得到大量完整或近乎完整的基因组序列，我们只要测得一小段蛋白质的序列，就可以通过寻找匹配的可读框来鉴定这个蛋白质的编码基因。

Edman 降解（Edman degradation）是化学反应，反应中多肽链的 N 端挨个释放氨基酸残基（图 21-23）。这种方法的关键点是一种叫作苯异硫氰酸酯（phenylisothiocyanate，PITC）的化学试剂能特异地修饰肽链的 N 端氨基酸的游离 α-氨基基团。用酸处理后，这一被修饰的末端氨基酸能从多肽中切离，在这个过程中要控制反应条件使其他氨基酸不受酸的破坏。通过分析高效液相色谱（HPLC）的洗脱峰可以鉴定末端氨基酸衍生物的性质（每种氨基酸都有一个特征性的滞留时间）。每一次肽链裂解都能产生一个带有游离 α-氨基基团的正常 N 端。因此，Edman 降解可以重复很多个循环，从而揭示蛋白质 N 端区段的序列。实际应用中，蛋白质的鉴定通常要进行 8～15 轮 Edman 降解。这些数量的循环通常足以特异性地鉴定一种蛋白质。

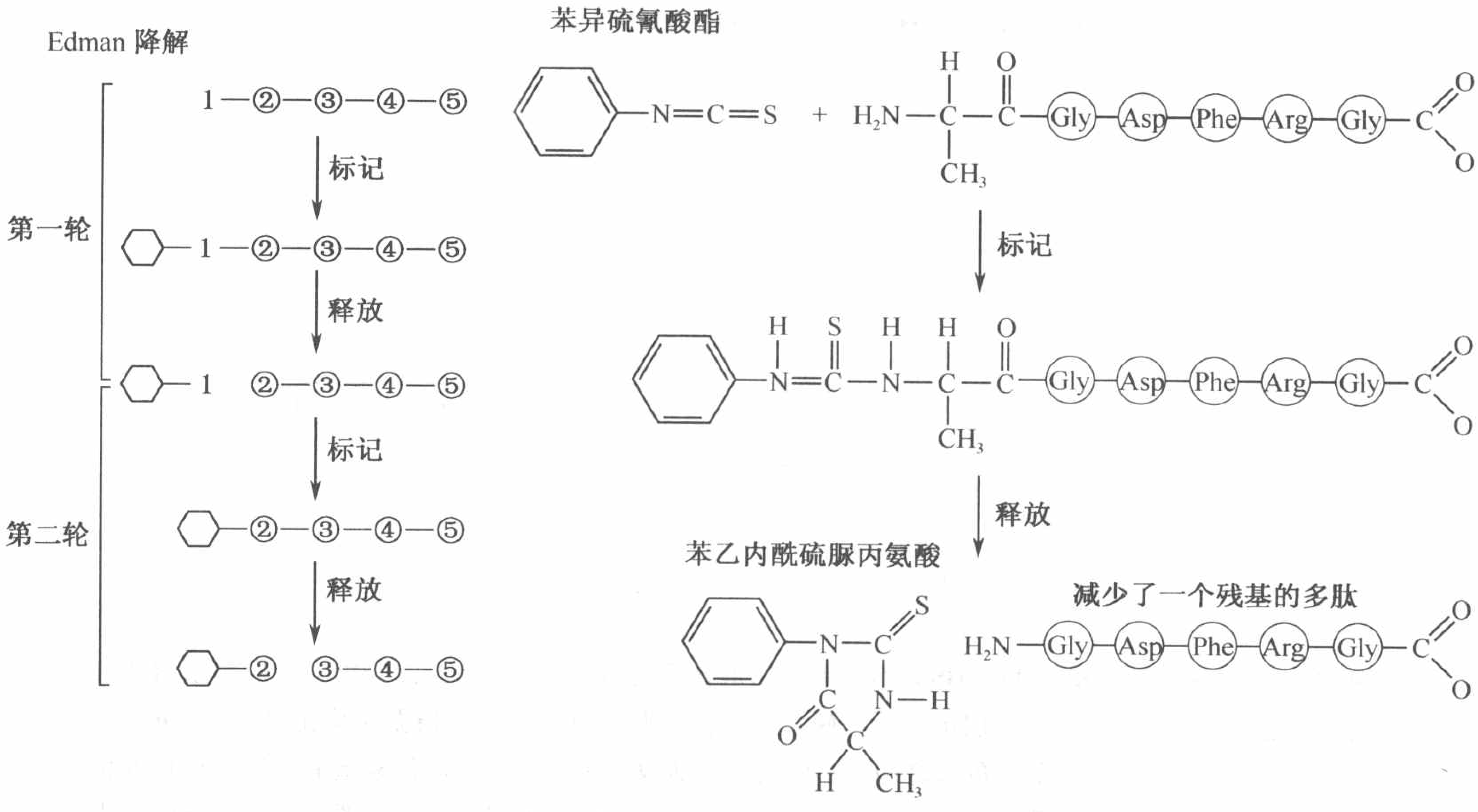


图 21-23　用 Edman 降解法进行蛋白质测序。 N 端残基可以被标记，并在不水解肽链的其余氨基酸的前提下切去这个被标记的残基。因此，每一轮反应能鉴定多肽序列中的一个残基。而这个残基代表肽链序列中的下一个残基。

自动 Edman 降解的 N 端测序已被广泛使用。然而，如果蛋白质的 N 端已被化学修

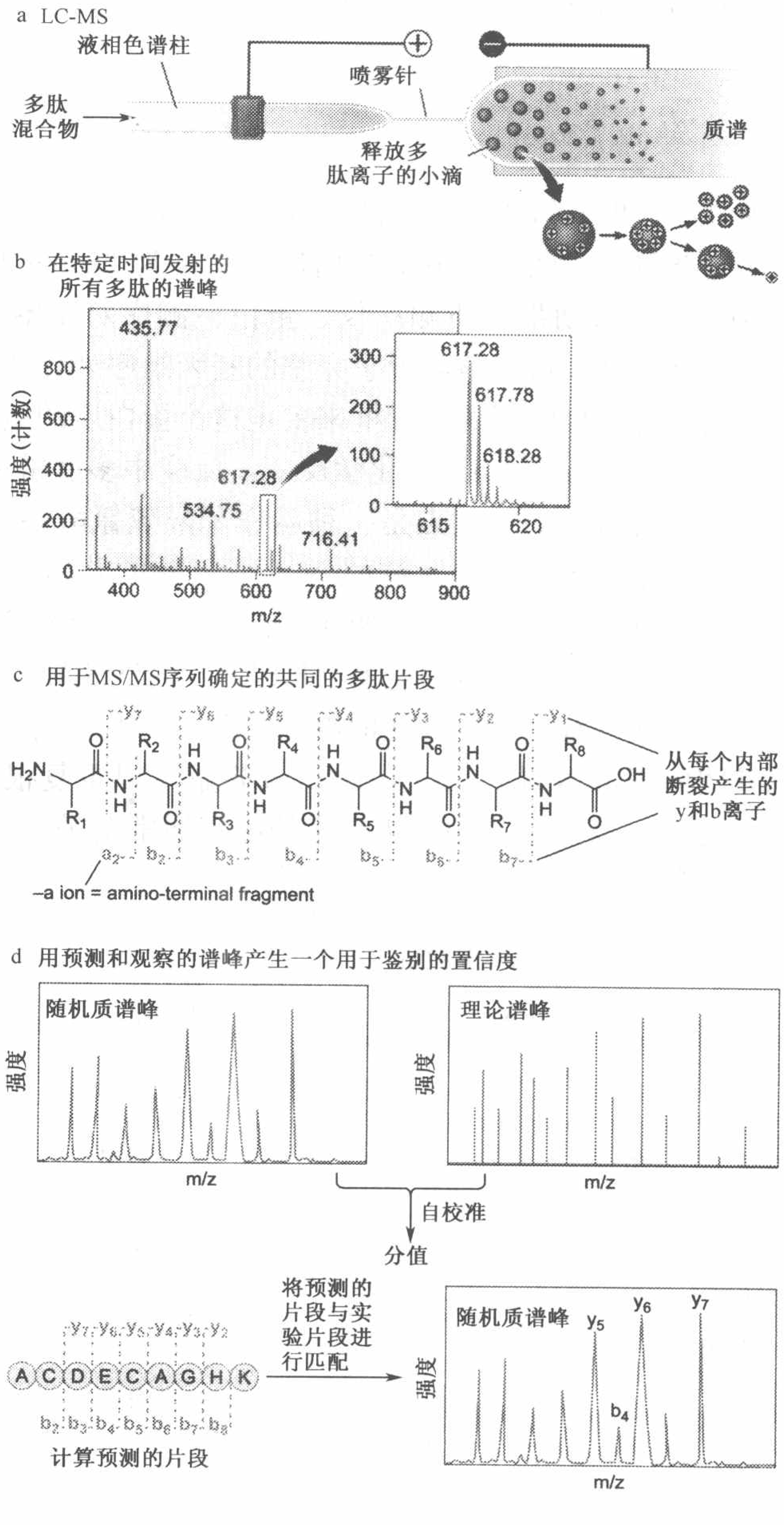


图 21-24 联用液相色谱和 MS/MS 法分析蛋白质混合物的组分。(a) 肽段混合物导入后接质谱装置的液相色谱。(b) 部分肽段从层析柱上洗脱，质谱仪对肽段进行分离，根据质荷比（m/z）将结果展示。选择相关肽段（相邻近的峰之间的不同是由于肽段上不同的原子同位素造成的）进行片段化并对所得的肽链片段用二级质谱进行分析。(c) 片段化后产生的不同的断点的图示。可能产生的小片段多肽包括 b 肽（氨基末端片段）、y 肽（羧基末端片段）、a_2 肽（最短的氨基末端片段）。(d) 将观察到的图谱与所有可能的理论图谱（根据来自生物的蛋白质序列）进行比对，理论图谱由分离此蛋白质群的生物体中编码这些蛋白质的氨基酸序列生成。一般来说，只有很少的一部分肽段可以精确鉴定。例如，异亮氨酸和亮氨酸的分子质量相同。然而，只要有 3 或 4 个准确鉴定的肽段，就足以进行蛋白质的鉴定。

饰（如被甲酰化或乙酰化），测序就比较困难。这种化学修饰所导致的封闭可能在体内发生，也可能在蛋白质分离过程中产生。当蛋白质的 N 端被封闭时，通常用蛋白水解酶消化，暴露出蛋白质的内部序列再进行测序。

随机质谱法（tandem mass spectrometry，MS/MS）也能用来进行蛋白质的测序。它能非常准确地测定微量样品的质量。其基本原理是物质在（真空）仪器中的行进与它的荷质比密切相关。对于小的生物大分子如肽段和小蛋白质来说，所测得的分子质量能精确到一个道尔顿。

用 MS/MS 进行蛋白质测序前，我们通常用特异性蛋白水解酶，如胰蛋白酶，将目的蛋白消化成小肽段（一般小于 20 个氨基酸）。质谱分析时肽段混合物中的各个肽段由于荷质比不同，能相互区分开来。将如此分离到的单个肽段打碎成其组成肽段，对这些组成肽段进行测序（图 21-24）。运用反卷积法分析这些序列数据能清楚地反应原始肽段的序列。和Edman降解一样，我们只要得到某种蛋白质的约 15 个氨基酸长的肽段序列，就能通过比较实测氨基酸序列和经 DNA 序列预测到的氨基酸序列而鉴定这种蛋白质。

MS/MS 使蛋白质测序和鉴定有了革命性的突破。它的特点是只需使用极其微量的样品，并能对复杂混合物中的各种蛋白质同时进行分析。

蛋白质组学

在全基因组水平对基因表达的测定为细胞活性状态提供了一个快照。但是这种方式不能检测其他重要的调控水平。实际上，基因的转录水平所能提供的信息只能大致的反映编码蛋白质表达水平。如果 mRNA 寿命很短或者很少被翻译，那么即便是高度表达的 mRNA 也只能产生少量的蛋白质。另外，很多蛋白质是通过翻译后修饰的方式显著地影响自身的活性，转录谱所提供的数据则根本不具备反映这种调节水平的功能。

全基因组序列的获得及高通量的蛋白质分离、鉴定方法的研究引入了蛋白质组学领域。蛋白质组学的研究目标是鉴别特定条件下细胞或组织产生的所有蛋白质（也就是蛋白质组），包括全套蛋白质的相对丰度、修饰以及与之相互作用的蛋白质。鉴于芯片使在全基因组水平研究转录谱或 DNA 含量研究成为可能，蛋白质组学的研究方法则着力于获得细胞内的所有蛋白质情况的类似快照。

联合使用液相色谱和质谱鉴定复杂提取物中的蛋白质

联合使用液相色谱和质谱（在本章前面的段落中已作过介绍），对鉴定复合体系中（如细胞粗提物）的所有蛋白质是一种有效的方法。理论上讲，人们可直接使用质谱技术分析出细胞提取物中的所有蛋白质，但实际上却很难做到，这是由于提取物中数目巨大的蛋白质产生的肽段超出可解析的能力。所以研究者们发展了有效的方法，在进行质谱分析之前，先通过两类液相色谱将肽段分离（LC-MS，图 21-25）。在这种方法中，首先用序列特异性的蛋白酶对细胞粗提物进行消化（如胰蛋白酶，可在精氨酸或赖氨酸残基后切开）生成多肽链，再使用离子交换层析（肽段与带电柱料发生离子相互作用分

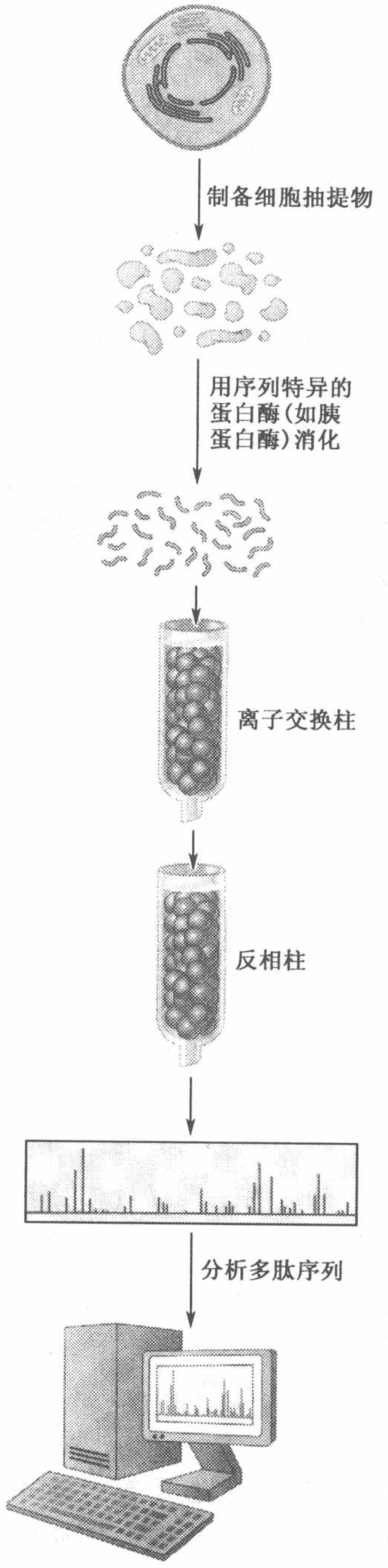


图 21-25　在液相色谱分离后进行质谱分析来分离蛋白质。这种方法的步骤如图所示，并在文中有详细记述。

离）和反相层析（肽段与柱料的疏水相互作用分离）将所得到的多肽进行分离。这一步骤使复杂的初始混合物简化为组分简单、容易区分和测序的二组混合物。接着对这些二级混合物进行串联质谱分析（MS/MS，前文已作讨论）鉴定多肽群中尽可能多的肽段。最后，在获得所研究生物体完整基因组序列和质谱分析得到的多肽序列的情况下，用生物信息学手段将每个多肽与基因组中蛋白质编码序列（基因）对应起来。

在实际操作中，这种方法只能检测复杂体系中某个亚体系的蛋白质，如从细胞中分离得到的蛋白质。常规分析只能检测到大约 1000 个不同的蛋白质。其他的分离方法和质谱仪灵敏度的提高，将来会进一步提高蛋白质谱的完整性。对鉴别细胞中蛋白质的存在情况来说，LC-MS 分析非常有效，但目前用此方法确定蛋白质的相对丰度还是比较困难的。为了解决这个问题，人们正在研发丰度定量的技术并已在某些情况下进行运用。

通过蛋白质组学比较鉴定细胞间的差异

虽然鉴定出细胞中的全部蛋白质具有深远的意义，但在多数情况下，知道两个不同细胞型之间蛋白质表达差异，或者同一细胞型在不同生长条件下的蛋白质表达差异，实际上是更有意义的。通过确定特定生理状态下的蛋白质组，就能够找到蛋白质表达水平的差异，进而确定那些可能造成不同细胞差异的蛋白质，为进一步的研究提供有价值的候选蛋白质。

通过对不同癌细胞的分析，可以了解比较蛋白质组学的价值。我们经常发现，虽然不同患者具有同一癌症的生理特征，但对同一化疗药物却表现出不同的反应。通过比较不同肿瘤样本在蛋白质组水平的差异，发现表观相似的细胞在其所表达的蛋白质上产生了很大差异。这些差异蛋白质在肿瘤分型中可以作为有价值的标记物。更重要的是，这些标记物可以为患者筛选有效的化学治疗方法提供依据。

质谱技术也可以检测蛋白质的修饰情况

鉴于蛋白质的修饰可以广泛影响其功能，研究人员也发展了多种用于鉴别细胞中蛋白质修饰的手段。特异的修饰作用通常改变蛋白质的活性或稳定性。如蛋白质的磷酸

化经常被用于调控其活性。磷酸化引起蛋白质构象的变化，进而影响功能（如许多蛋白激酶只有在磷酸化状态下才具有生物学活性）。另外，磷酸基团的结合也会使蛋白质产生新的蛋白质结合位点，导致新的蛋白复合体的形成。其他的蛋白质修饰还包括甲基化、乙酰化和泛素化。泛素化是蛋白质的赖氨酸残基和一个包含 76 个氨基酸残基的泛素蛋白通过类肽键连接。一个蛋白质被多个泛素蛋白质修饰，将触发蛋白质的降解。

每一种修饰作用都会使蛋白质的分子量发生不同的变化。这种变化可以用质谱检测，并发展出检测细胞中只带有某种特定修饰的蛋白质谱的方法。例如，细胞内所有的磷酸化蛋白质称为“磷酸蛋白质组”。已经建立的鉴定亚群蛋白质的方法一般采用能和发生特定修饰的蛋白质结合的亲和树脂。例如，含有固定化三价铁离子（Fe^{3+}）的树脂（也称固定化金属亲和层析［IMAC］）可以特异性结合磷酸化肽段，细胞粗提物中的肽段混合物可以与它共同孵育，使磷酸化肽段富集。得到的肽段就可以借助 LC-MS 方法进行蛋白质磷酸化和磷酸化位点分析。所获得的信息对测定激酶修饰蛋白质和制备不能修饰的激酶突变体以检测修饰的重要性是非常重要的。

蛋白质-蛋白质相互作用可以产生蛋白质功能的信息

蛋白质组学也包括细胞中蛋白质的相互作用，我们称其为互作组（interactomes）。一个细胞中完整的互作组应该反映细胞中所有的蛋白质相互作用。在整个互作组中，一般认为主要是蛋白质之间的关联在起作用。通过这样的相互作用，我们可以推测出一个蛋白质在怎样的细胞内代谢过程中起作用。同一个蛋白复合体中的蛋白质几乎总是在同一个细胞代谢过程中出现。

酵母双杂实验是一种鉴定蛋白质相互作用的方法（框 17-1），这项技术把感兴趣的蛋白质作为“诱饵”，而在待检测的蛋白质库中可能存在被捕捉的“猎物”。另一种方法是，用亲和树脂或免疫共沉淀对目的蛋白及相关蛋白质进行快速纯化。可以通过 LS-MS 方法在得到的蛋白质混合物种鉴定相关蛋白质。在一个细胞所有蛋白质中重复进行这样的鉴定，就会得到这个细胞中蛋白质之间的互作图。

第二种方法已经在酿酒酵母中得到了应用。通过亲和层析已经纯化了 6000 多个酵母蛋白质（每一个蛋白质对应的基因都进行了遗传改造，或者将一个标签融合在 C 端，该标签可以和两种亲和树脂结合），用质谱鉴定与标签蛋白质共同纯化的结合蛋白质。通过对这些数据的分析，鉴定了数百个酵母蛋白复合体，虽然这些复合体有许多是已知的，但也有一些是新发现的。这项研究的有效性，可以通过对已知的蛋白质复合体的检测进行验证（如 RNA 聚合酶Ⅱ；图 21-26）。

核酸-蛋白质相互作用

下面我们把关注的重点放到研究核酸和蛋白质相互作用的各种方法上。他们之间的相互作用对决定本书所描述的许多事件的特异性和精确性来说，都至关重要。在转录、重组、DNA 复制、DNA 修复、mRNA 剪切或翻译中，介导这些事件的蛋白质必须识

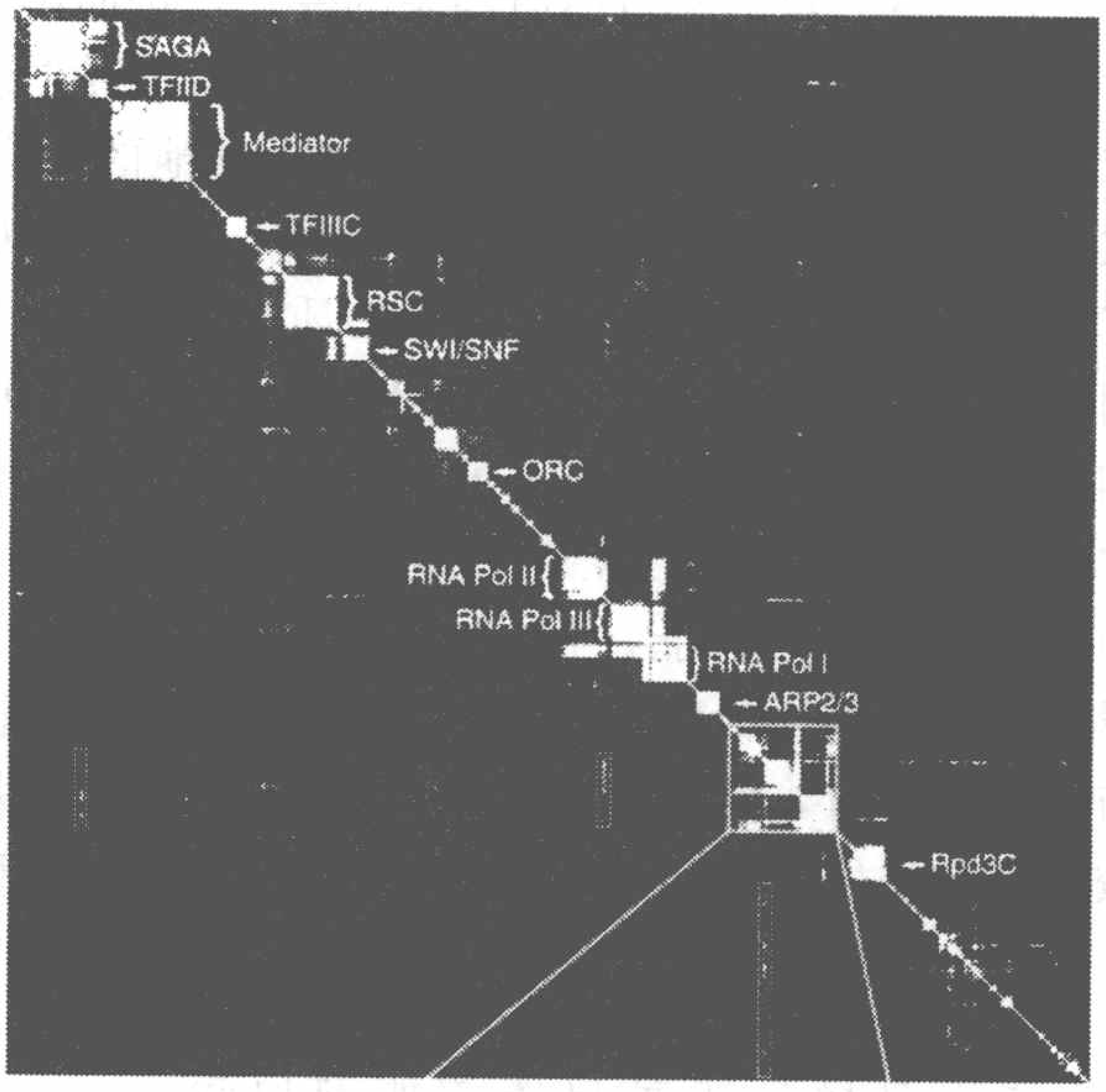


图 21-26 酿酒酵母的物理互作图。图中所示利用亲和纯化/质谱研究酿酒酵母中所有蛋白质后，得到的分析结果。全图由一系列的框图构成，这些框图是指能和给定蛋白质共沉淀的蛋白质。如果一个蛋白质可以和标记蛋白质共沉淀，框用黄色标记，如果不能发生共沉淀，框就是黑色的。此图中，同一个蛋白复合体中的蛋白质会在纵轴和横轴上表现出成簇聚类现象；因此，复合体都出现在对角线上。在这幅图中，只有部分蛋白质复合体（很多复合体在本书的其他部分已做探讨）被展示出来。框内的区域是像在第五部分第一项的第一张图一样被放大展示的。（经许可影印自 Collins S. R. et al，2007. *Mol. Cell. Protem.* 6：439-450，Fig. 3b. © American Society for Biochemistry and Molecular Biology.）

别核酸结构或者序列，才能确保生命活动在细胞的特定时空发生。

基于核酸-蛋白质相互作用的重要性，研究者开发了多种有效的用于体内和体外检测的方法。在接下来的部分，我们将介绍几种分析方法，并探讨它们的优势和不足。

蛋白质的结合改变 DNA 的电泳迁移速率

正如电泳迁移速率可以确定 DNA、RNA 或蛋白质的相对大小，它也可以用于检测蛋白质-DNA 之间的相互作用。如果一个 DNA 分子结合了蛋白质，同未结合蛋白质的 DNA 相比，DNA-蛋白质复合体在凝胶中的迁移速率变慢。这就是检测特定 DNA 结合活动的理论基础。一般的实验流程为，对含有感兴趣结合位点的短 DNA 片段进行放射性标记，这样就可以通过聚丙烯酰胺凝胶电泳和放射自显影检测到微量的样品。也可以使用荧光标记，但是荧光染料不能处在干扰 DNA 结合的位置（参见框 8-1 图 1b）。接下来，以标记 DNA 为探针与待测蛋白质混合，使用非变性凝胶电泳分离。如果发生结合，则蛋白质-DNA 复合体的迁移速度就比较慢，标记 DNA 在凝胶上的位置会发生变化（图 21-27）。这就是电泳迁移改变检测（EMSA），或者更简单地说成是条带或凝胶迁移检测。我们虽然只介绍了用 EMSA 检测蛋白质和双链 DNA 的相互作用，EMSA

也可以用于检测蛋白质与单链 DNA 或者 RNA 的结合情况。

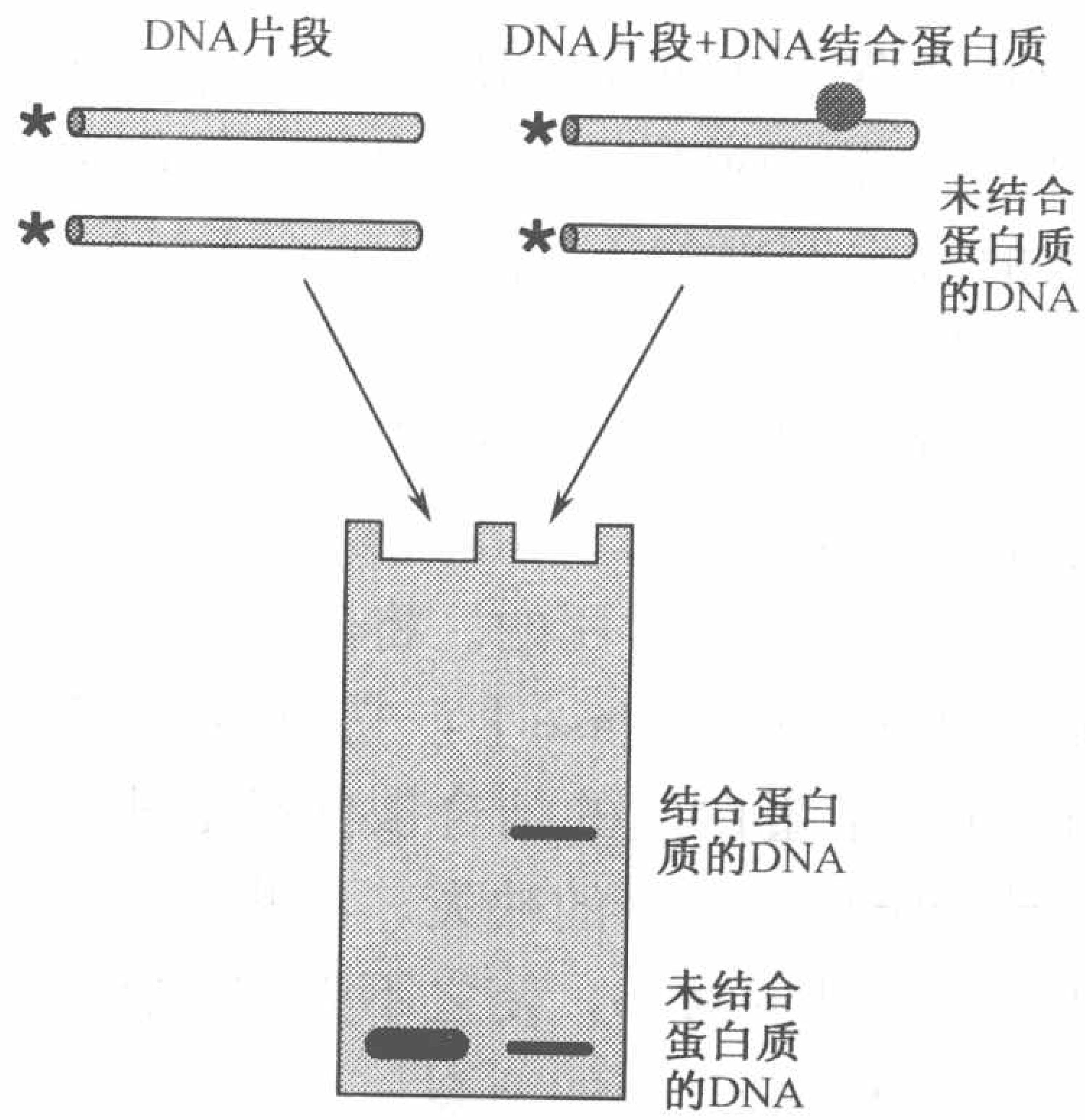


图 21-27 电泳位移分析。位移分析的基本原理见示意图。蛋白质与放射性标记的 DNA 探针混合，DNA 探针具有该蛋白质的结合位点。聚丙烯酰胺凝胶电泳分离混合物，再通过放射自显影显色。没有结合蛋白质的 DNA 在凝胶上呈现一条带，与 DNA 片段的大小对应（左边泳道）。混合物中，由于只有一部分 DNA 分子（不是所有的都与蛋白质结合）结合蛋白质分子，所以右边泳道出现两条带，一条对应未结合蛋白质的分子，另一条对应与蛋白质结合的分子。

EMSA 也可以检测多个蛋白质和同一个 DNA 的结合。这些相互作用可能都是由于序列特异的 DNA 结合造成的。或者是，在一个 DNA 序列特异的结合发生后，另一个蛋白质可以同第一个 DNA 结合蛋白质结合。不管在什么情况，只要有新的蛋白质结合上去，DNA 的凝胶迁移速度都会进一步降低。EMSA 是一种有力的工具，它可以鉴定一系列蛋白质是如何独立地与同一个 DNA 相互作用的。蛋白质分子质量不同，影响 DNA 迁移速率的效果也不同（蛋白质分子质量越大，迁移速率较慢），所以不同蛋白质结合同一种 DNA 分子也是可以区分的。如果不同的蛋白质的结合产生相同的迁移，我们可以采用另外的方法确定哪一个蛋白质发生了结合。当蛋白质与 DNA 结合时，蛋白质对应的抗体的加入可引起超迁移作用。因此，通过加入可能与 DNA 结合下蛋白质抗体，可鉴定在这一 DNA-蛋白复合体中是否存在这一蛋白质。

EMSA 的缺点在于无法确定 DNA 片段中与蛋白质特定结合的核酸序列。两种办法可以解决这个问题。第一种方法是在结合探针 DNA 之前，用一些短的双链 DNA 寡聚物与蛋白质共同孵育，如果寡聚物中存在待测蛋白质的结合位点，那么孵育后，蛋白质就不会与探针 DNA 结合。另一种方法是应用定点突改变探针 DNA 序列，从而推测出它对蛋白质结合的影响。这些方法可以在蛋白质结合位点未知的情况下进行，在多数情况下，加入寡聚物或改造探针 DNA 的序列对简化待测序列的选择是很有帮助的。

蛋白质的结合可以保护 DNA 不被核酸酶降解或化学修饰

如何更加快捷地确定 DNA 序列中的蛋白质结合位点，如操纵子位点？已经有一系列的手段可用来确定结合位点，以及确定蛋白质相关联的化学修饰位点（甲基化、氨基化或者磷酸化）。这些方法的基本原理如下：如果一个 DNA 分子只在其每股单链的末端标记一个同位素，那么单链上从任何位置断开，就会使被标记片段分子质量发生不同程度的减小。通过高分辨力变性聚丙烯酰胺凝胶电泳，以及检测标记单链 DNA 片段可确定这些片段的大小。这就是通常所说的 DNA 足迹法。

核酸酶保护足迹法是这些方法中最常用的。将蛋白质与末端标记的 DNA 共同孵育后，产物直接与 DNA 酶反应（常用 DNase Ⅰ，它可以切割双链 DNA 中的一条链）。DNA 序列上存在的酶切位点由于蛋白质的结合被保护起来，不能被切开（图 21-28），所得的“足迹”显现的就是特定大小条带的缺失。与此相关的化学保护足迹法则是基于结合的蛋白质可以保护结合位点处的碱基，使之不受碱基特异的化学试剂的影响，这些试剂（在进一步反应后）能导致核酸骨架的断裂。在两种方法中，重要的是要保持核酸酶切割位点或者化学修饰位点数目大约为每个 DNA 探针一个。这是因为只有那些距离标记 DNA 末端最近的酶切位点可以被凝胶电泳和标记 DNA 检测到。

改变足迹法实验的前两步的顺序，建立了第三种方法——化学干扰足迹法。这种方法可以确定什么样的 DNA 结构对蛋白质结合是必要的。在与蛋白质孵育之前，对每一个 DNA 分子进行化学改造。改造后的 DNA 分子与 DNA 结合蛋白质孵育，然后分离 DNA-蛋白质复合体。从总 DNA 中将结合了蛋白质的 DNA 片段分离最常用的方法是 EMSA。在 EMSA 凝胶中检测标记 DNA，很容易区分迁移发生变化（结合）和没有发生变化（未结合）的 DNA。如果某一个位点的修饰不能阻止蛋白质的结合，那么从复合体中分离得到的 DNA 就带有那样的修饰。反之，如果修饰阻止了蛋白质识别 DNA，在蛋白质结合样品不会找到在那个位点被修饰的 DNA。就像化学保护分析方法一样，也可以用试剂切开 DNA 的化学修饰位点，来确定化学修饰位点。

用于化学修饰的试剂可以非常特异地检测 DNA。例如，化学试剂乙基亚硝基尿（ENU）可以特异地修饰 DNA 骨架上的磷酸基团。其他试剂也可以特异地修饰 DNA 大沟或小沟上特定的碱基。通过使用不同化学试剂，可以精确了解特定的蛋白质与 DNA 核糖-磷酸骨架上特定碱基，或者磷酸基的相互作用。

足迹法是一种快速鉴定 DNA 序列中蛋白质结合位点的有效方法，但是，该方法在分析中针对的是蛋白质与 DNA 强力的结合。总体来说，为确保 DNA 足迹法的有效性，必须有超过 90%的 DNA 探针与蛋白质结合。足迹法检测的是信号缺失（因为 DNA 的酶切保护和修饰作用），而不是检测新条带的产生，所以高水平的结合是很必要的。这与高灵敏度的 EMSA 是不同的，在 EMSA 分析中，如果蛋白质与标记 DNA 结合，会在凝胶中出现新的条带，如果没有结合，则没有新条带的出现。

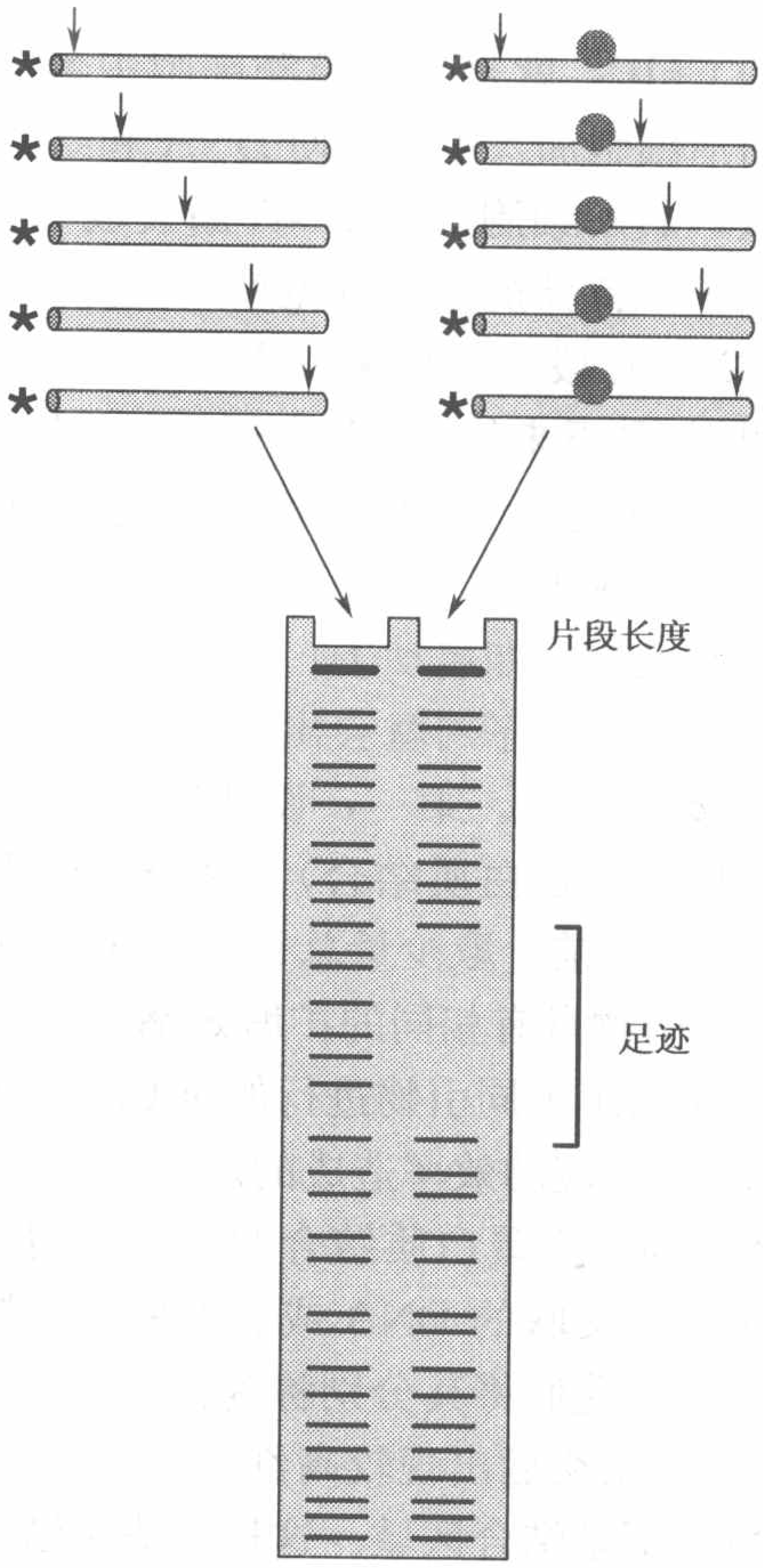


图 21-28 核酸酶保护足迹法。“*”号表示放射性标记的 DNA 片段末端，箭头表示 DNase 酶切位点，红圈表示 Lac 抑制因子结合在操纵子上。左侧，DNase 随机切割 DNA 后，根据大小通过电泳分离。右侧，DNA 分子先结合抑制因子，再用 DNase 处理。“足迹”左右侧泳道中显示出来。这些条带对应的是 DNase 降解的游离的 DNA，而非结合了抑制因子的 DNA。在后一种情况下，DNA 上的切割位点在操纵子序列中，被抑制因子所覆盖，使 DNase 不能发挥作用。

染色质免疫共沉淀检测细胞中蛋白质与 DNA 的关联

虽然蛋白质与 DNA 相互作用的体外分析，可以提供许多重要信息，但确定蛋白质与特定 DNA 是否在活细胞中存在特定关联更有意义。对于特定的 DNA 结合蛋白，在整个基因组中会有许多潜在的结合位点。尽管如此，在很多情况下只有少数位点能与蛋白质结合。与其他蛋白质的结合有时会抑制这个蛋白质与潜在 DNA 结合位点的关联。有时，稳定的关联还需要与相邻蛋白质的结合。不管在什么情况下，知道细胞中一个蛋白质（如转录调控因子）是否与一个特定的 DNA 位点（如特定启动子区域）发生结

合，都会为在这个位点发生调控事件（如转录激活）提供有力证据。

染色质免疫共沉淀，即 ChIP，就是一种检测细胞中蛋白质-DNA 相互作用的常用手段。其过程大致如下：向活细胞加入甲醛，使 DNA 和结合蛋白质以及与其他蛋白质紧密结合的蛋白质发生交联，裂解细胞，将 DNA 断裂成 200～300bp 的片段。用针对目标蛋白质（如转录调控因子）特异抗体进行免疫共沉淀反应（IP），可从细胞裂解物中将与蛋白质结合的特定 DNA 片段分离出来。完成免疫共沉淀以后，再解开 DNA 和蛋白质之间的关联，就可以对 IP 下来的 DNA 序列进行分析。

ChIP 技术的关键环节在于确定特定 DNA 区域是否与蛋白质结合而出现在沉淀产物中。回答这个问题，可以有两种方法。第一种方法是设计与结合区域匹配的引物进行 PCR 反应，以检验 DNA 特定区域（如启动子）是否与目的蛋白质相连。如果在交联过程中蛋白质与 DNA 相互结合，那么序列就会出现在沉淀中，可以通过 PCR 反应扩增检测。这种方法需要二个重要的对照。第一，设计针对其他的 DNA 区域的 PCR 引物（该区域已知或者预计不能和目的蛋白质结合）。这个反应中，不会出现 DNA 的扩增（图 21-29）。第二，进行共沉淀之前，取少量总 DNA，用检测引物和对照引物对未分离的 DNA 样品扩增。如果二种引物具有相同的扩增效率，针对二种靶基因的扩增效率也应该相同。这个对照保证了用两组不同引物进行的对 ChIPDNA 扩增产生的差异是由于丰度差异造成的，而不是 PCR 反应的效率造成的。

第二种鉴定细胞内 DNA 与特定蛋白质结合的方法是覆瓦式 DNA 芯片技术。在这种方法中，分别将那些与蛋白质交联的 DNA 或总 DNA 分离出来，用两种不同的荧光染料进行标记（对于这个例子，我们将其分别称为红色和绿色）。然后把这两种不同的 DNA 混合，与芯片杂交。那些杂交后出现较高红∶绿比值的区域可以确定为蛋白质的结合区域，而低比值的区域就不是结合区域。这种方法对结合位点未知的序列检测很有效，而且能在这个基因组范围内同时进行检测。

虽然芯片技术非常有效，而且应用广泛，但我们应该对其在检测方面的局限性有所了解。首先，与 EMSA 相似，ChIP 的分辨率是有限的。ChIP 并不能确切了解与蛋白质相互作用的特定 DNA 片段，而只能是结合位点所在的 200～300bp 的序列。因此，ChIP 可以检测到调控蛋白在一个特定基因上游的区域发生结合，但不能显示出在基因上游的具体什么位置与蛋白质结合。和 EMSA 一样，定点突变对确认蛋白质是否特异地结合于某一位点是必要的。其次，只有那些获得对应的抗体目标蛋白质，才可以用 ChIP 进行分析。更重要的是，DNA 发生交联后（或者与其他蛋白质发生作用以后），待沉淀的蛋白质表位（抗体识别的蛋白质特定区域）仍然是暴露的，才能使用 ChIP 检测。进一步讲，如果一个蛋白质不能在某一个环境或生理条件下被检测到，而在其他环境下能检测到，一种直接的解释就是该蛋白质只有在环境变化时，才会与特定 DNA 区域结合。当然也可以有另外的解释，就是虽然这个待测蛋白质始终与 DNA 序列结合，但是其他结合蛋白质掩盖了它的表位，只有在某些特定条件下表位才能暴露。

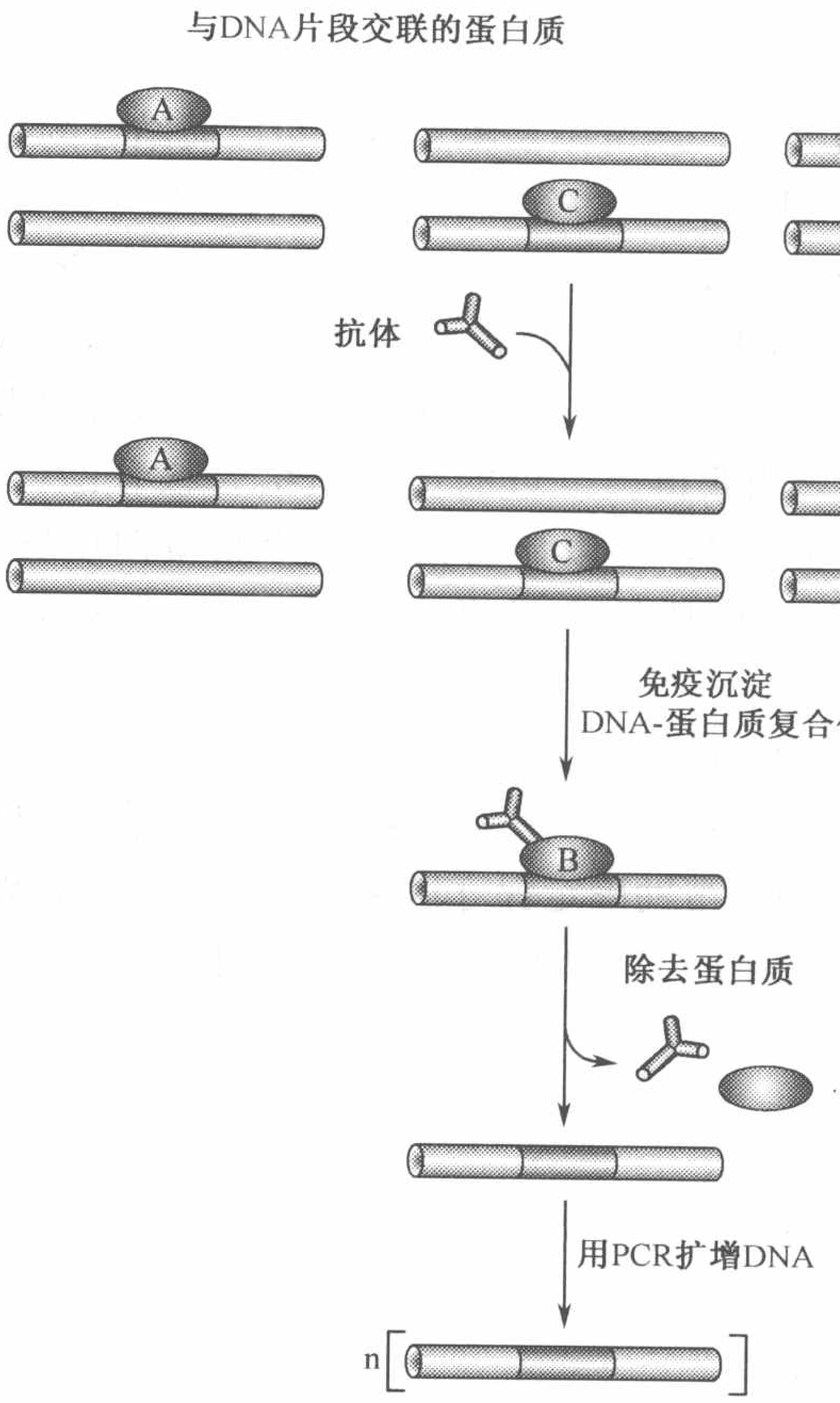


图 21-29 染色质免疫共沉淀（ChIP）。

可用体外选择鉴定蛋白质的 DNA 或 RNA 结合位点

随着越来越多 DNA 结合蛋白被鉴定和理解，确定与特定序列 DNA 结合的蛋白质的序列特异基序（motif）（例如，helix-turn-helix 基序）也相对容易了。尽管如此，对这些结合蛋白结构域的了解还存在着很多不足，如何根据氨基酸一级结构确定与什么序列的 DNA 结合，依然难以回答。而通常获得这些信息对我们确定可能的调控区以做进一步的分析研究是非常重要的。

如何鉴别与特定蛋白质结合的 DNA 序列？体外选择，或称 SELEX（systematic evolution of ligands by exponential enrichment，指数富集配体的系统进化），是目前一种有效的手段。该方法利用蛋白质序列特异性，在不同的寡聚核苷酸文库中筛查。然后分析富集的 DNA，就能确定与蛋白质紧密结合的 DNA 序列。

SELEX 的第一步是利用 DNA 化学合成技术（本章第一部分已做介绍）合成一个巨大的单链 DNA 文库。重要的是这些单链 DNA 中部的 10～12 个碱基是随机的（合成过程到达这些位点时加入所有四种核苷酸前体的混合物）。随机区域的两侧是特定的

DNA 序列，文库合成后，用短的与特定区 3′端匹配的引物进行延伸，就能将随机单链 DNA 文库变成双链 DNA 文库。

用与我们曾讨论过的类似方法，就可以富集那些与蛋白质结合的寡聚核苷酸。蛋白质与寡聚核苷酸共同孵育后，用 EMSA 法分离。在电泳位置发生迁移的复合体中，富集着与蛋白质紧密结合的 DNA。或者，如果有待测蛋白质的对应抗体，可以使用与 ChIP 类似的免疫共沉淀从未结合 DNA 中分离出蛋白质和结合的 DNA。不管采取什么样的富集机制，用 PCR 就可以扩增结合的 DNA（用与寡核苷酸特定末端区域配对的寡聚核苷酸链）。因为从文库中只能获得很少量的蛋白质结合核苷酸，所以扩增是必须的。重复结合、富集和扩增的步骤，可以显著富集与待测蛋白质紧密结合的核酸序列（图 21-30）。通常，找到与蛋白质结合最紧密的 DNA 序列只需三到五轮这样的循环。

测定富集的 DNA 序列，可以确定蛋白质结合 DNA 的序列特异性。一般来说，随机区域的 DNA 序列只有一部分是保守的，因为大多数 DNA 结合蛋白识别不多于 6 或 7 个核苷酸，计算机分析通常被用来帮助确定保守区域的序列。最后，碱基序列保守性可以用序列 logo 图表示，在 logo 图中，字母 G、A、C 或 T 的大小表示碱基在富集的寡聚核苷酸序列中出现的频率（图 21-31）。

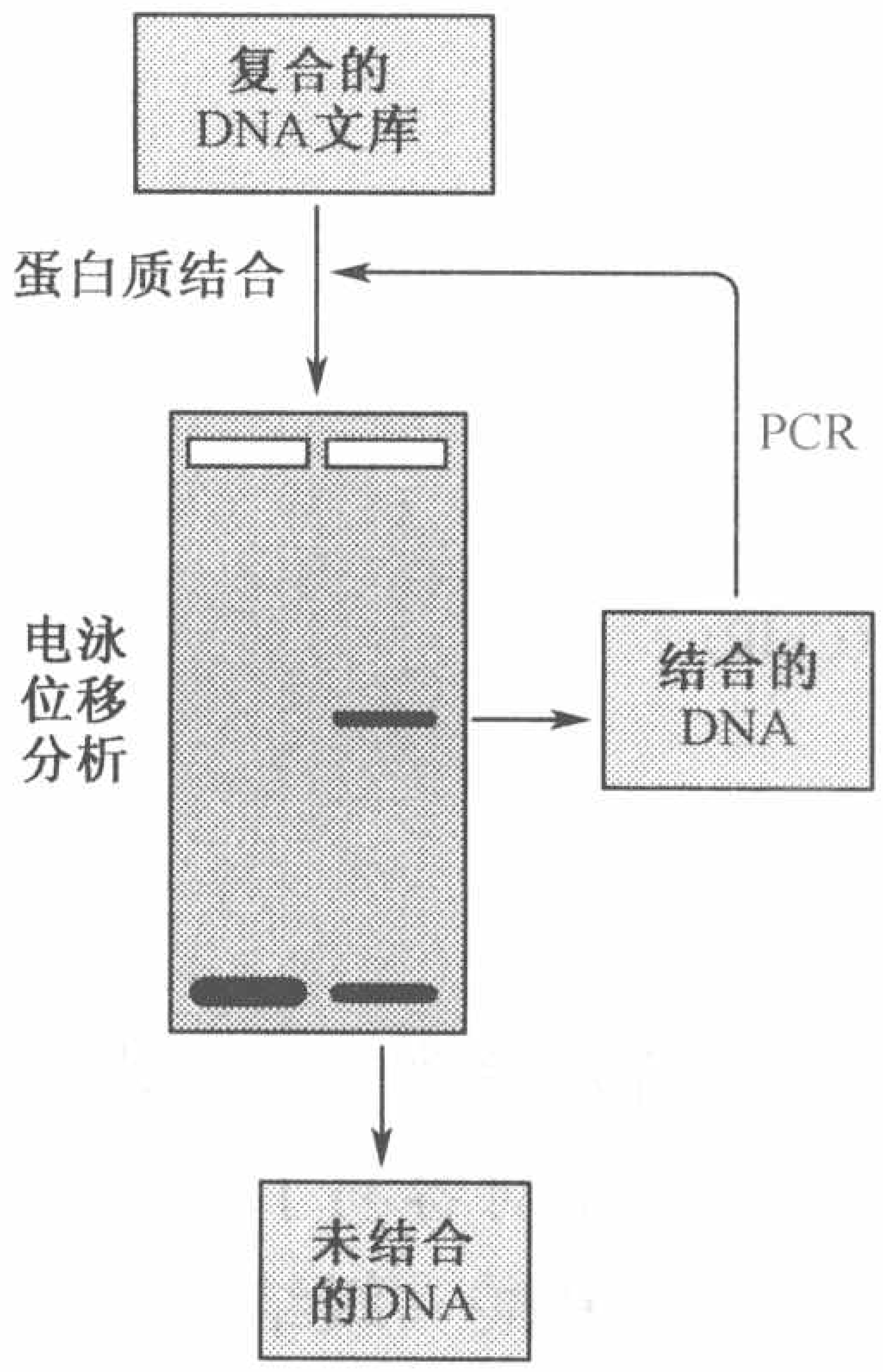


图 21-30 体外选择示意图。一个复合的 DNA 文库（库中 DNA 分子中部 10～12 个碱基是随机的）与待测蛋白质结合。与蛋白质结合的 DNA 分子通过 EMSA 与未结合的 DNA 相分离。结合的 DNA 可以从凝胶中洗脱下来，用与随机区域相邻的保守区域匹配的引物进行 PCR。经过二到五轮的结合富集循环，最后确定具有最高亲和度的序列。

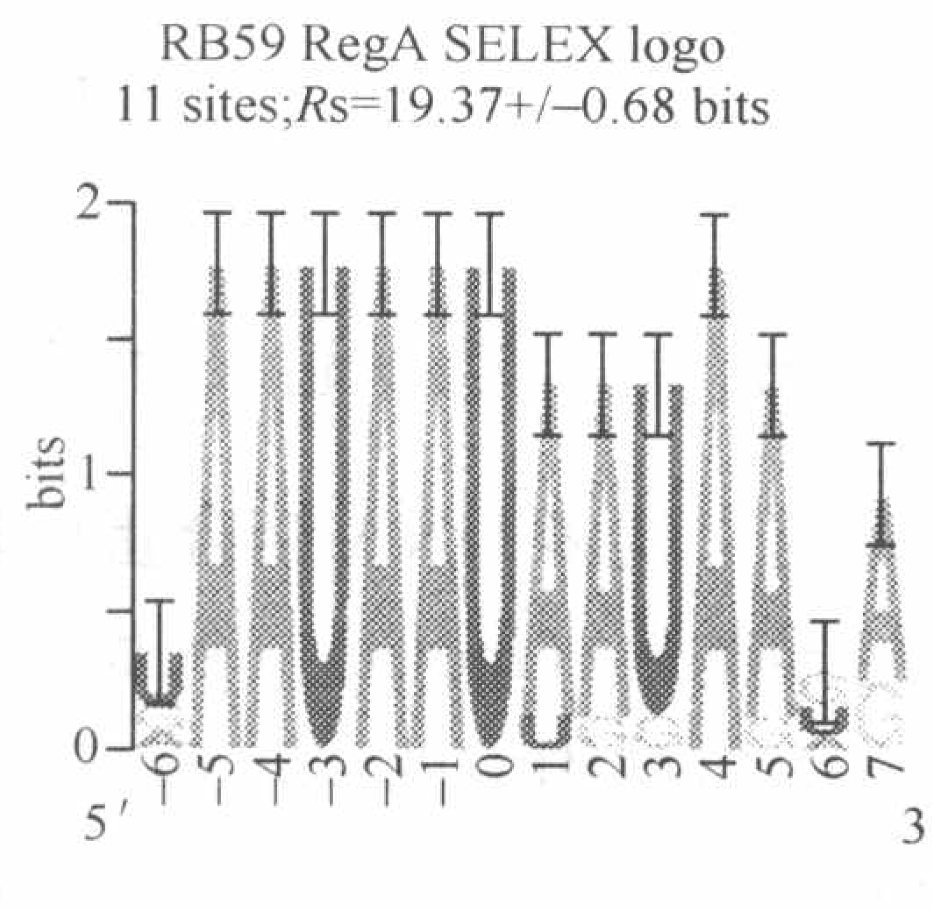


图 21-31 SELEX 序列 logo 图。体外筛选用于分离与翻译抑制蛋白 RB69 RegA 结合的 RNA。图示为获得的序列的 logo 图。字母高度与每个碱基在该位点出现的频率成正比，出现频率最高的碱基在最上面。(经允许影印自 Dean T. R. et al. 2005. *Virology* 366：26-36，Fig. 4a. © Elsevier.)

（朱金桂 译 刘国振 朱金桂 校）

参考文献

书

Griffiths A.J.F., Gelbart W.M., Lewontin R.C., and Miller J.H. 2002. *Modern genetic analysis*, 2nd ed. W.H. Freeman, New York.

Hartwell L., Hood L., Goldberg M.L., Reynolds A.E., Silver L.M., and Veres R.C. 2003. *Genetics: From genes to genomes*, 2nd ed. McGraw-Hill, New York.

Sambrook J. and Russell D.W. 2001. *Molecular cloning: A laboratory manual*, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Snustad D.P. and Simmons M.J. 2002. *Principles of genetics*, 3rd ed. Wiley, New York.

基因组分析

Human genome. 2001. *Nature* **409:** 813–960.

Human genome. 2001. *Science* **291:** 1145–1434.

International Human Genome Sequencing Consortium. 2004. Finishing the euchromatic sequence of the human genome. *Nature* **431:** 931–945.

Mouse genome. 2002. *Nature* **420:** 509–590.

蛋白质组分析

Yates J.R., 3rd, Gilchrist A., Howell K.E., and Bergeron J.J. 2005. Proteomics of organelles and large cellular structures. *Nat. Rev. Mol. Cell. Biol.* **6:** 702–714.

第22章 模式生物

本章概要

- 噬菌体
- 细菌
- 酿酒酵母
- 拟南芥
- 秀丽新小杆线虫
- 黑腹果蝇
- 小鼠

分子生物学中有一个著名的说法：基础问题可以在最简单和最容易获得的系统中得以回答。因此，长期以来，分子生物学家把精力集中在少数几种所谓的模式生物上。其中，重要的模式生物，以复杂程度增高为序有：大肠杆菌（*Escherichia coli*）和它的噬菌体——T噬菌体和λ噬菌体，酿酒酵母（*Saccharomyces cerevisiae*），拟南芥（*Arabidopsis*），秀丽新小杆线虫（*Caenorhabditis elegans*），黑腹果蝇（*Drosophila melanogaster*）和小鼠（*Mus musculus*）。

模式生物的共同点是什么呢？它们的一个重要特征是有传统遗传学和分子遗传学的强有力工具可资利用，使从遗传学角度操作和研究这些生物成为可能；第二点是每种模式生物都吸引了许多研究者。同一种生物的研究者之间共享观点、实验方法、实验系统和生物品系，会促进研究的快速发展。

例如，从20世纪40年代开始，以M. Delruck、S. Luria和A. D. Hershey为首的一群科学家，每年夏季都在纽约冷泉港实验室（Cold Spring Harbor Laboratory）研究大肠杆菌T噬菌体的增殖。这个小组，称为噬菌体小组（phage group），在创建分子生物学领域时发挥了重要作用。噬菌体小组的许多成员是物理学家，被噬菌体的研究所吸引。这不仅仅因为噬菌体相对简单，而且，每次实验可以研究大量噬菌体，从而产生定量的以及有显著的统计学意义的结果。到20世纪50年代后期，冷泉港实验室每年开设一次噬菌体课程，越来越多的研究者来学习这个新系统。这是一个典型的实例，表明集中研究相同的模式生物，肯定比分散研究多种不同的生物的进展更快。

选择哪一种模式生物取决于要回答什么问题。例如，研究分子生物学的基本问题，用简单的单细胞生物或病毒通常更方便些。这些生物可以快速、大量地生长，通常可把遗传学和生物化学的研究方法结合起来；而其他问题，如有关发育的问题，通常只能用

更复杂的模式生物来解决。

众所周知，T 噬菌体（如大家最熟悉的成员，T4）是一个解决基因和信息传递本质的理想系统；另一方面，因为有高效的适合遗传分析的交配系统，酵母则成为揭示真核细胞本质的首选系统。从真菌到高等细胞的进化保守性意味着在酵母中的发现在人类中也是常见的。线虫和果蝇也提供了很好的遗传系统，用来解决那些在较低等的生物中不能有效解决的问题，如发育和行为；还有一种模式生物小鼠，尽管它不如线虫和果蝇容易研究，但因为是哺乳动物，所以是了解人类生物学和人类疾病的最好的模式系统。

本章将描述一些最常用的实验生物，并逐一解释它们作为模式系统的主要特性和优点；另外还将介绍研究每种生物的实验系统以及已经研究的一些生物学问题，但不会对所有在分子生物学上有重要影响的模式生物进行详细解释。

噬菌体

噬菌体（及其他病毒，广义而言）提供了一个研究生命基本过程的最简单的系统。病毒基因组一般很小，只在侵入宿主细胞（对噬菌体而言，就是细菌）后才复制，所编码的基因才表达。在感染过程中基因组也会发生重组。

由于这个系统相对简单，在分子生物学的早期，噬菌体被广泛利用——事实上，噬菌体对分子生物学的发展至关重要。即便是现在，在研究 DNA 复制、基因表达和重组的基本机制时，它们仍然是一个可以利用的系统。另外，在重组 DNA 技术中（第 21 章），它们也是非常重要的载体，并且经常用来检测各种化合物引起突变的能力。

噬菌体通常由一个基因组（DNA 或 RNA，前者更为常见）被一层蛋白质亚基包裹起来而组成。蛋白质亚基形成一个头部结构，基因组就包在里面。有些蛋白质亚基形成尾部结构，尾部使噬菌体附着于细菌宿主细胞外面，噬菌体的基因组通过尾部注入到细胞里。其独特性在于：每一种噬菌体只附着于一种特定的细胞表面分子（通常是一个蛋白质），只有带有某种“受体”的细胞才会被一种特定的噬菌体所感染。

噬菌体有两种基本类型——裂解型（lytic）和温和型（temperate）。前者如 T 噬菌体，只进行裂解生长。如图 22-1 所示，当噬菌体感染一个细菌细胞，它的 DNA 得以复制，其基因组产生许多拷贝（多达几百个拷贝），并且表达编码新的外壳蛋白的基因。这些事件高度协调，以保证新的噬菌体颗粒在宿主细胞裂解释放它们之前构建起来。子代噬菌体又可以去感染其他宿主细胞。

温和型噬菌体（如 λ 噬菌体）也可以以裂解的方式复制，但它们还有另外一种发育途径，即溶原（lysogeny）（图 22-2）。处于溶原状态时，噬菌体基因组不是自己复制，而是整合到细菌的基因组里，并且外壳蛋白基因不表达。这种整合的抑制状态的噬菌体被称为原噬菌体（prophage）。原噬菌体作为细菌染色体的一部分在细胞分裂时被动复制，因此两个子细胞都是溶原性细菌。溶原状态可以如此保持几代，但同时也可以在任何时候转为裂解生长。这种从溶原到裂解途径的转变，叫做诱导（induction），涉及原噬菌体 DNA 从细菌基因组剪切出来、复制以及用于制造外壳蛋白和调节裂解生长的基因的活化（图 16-24）。

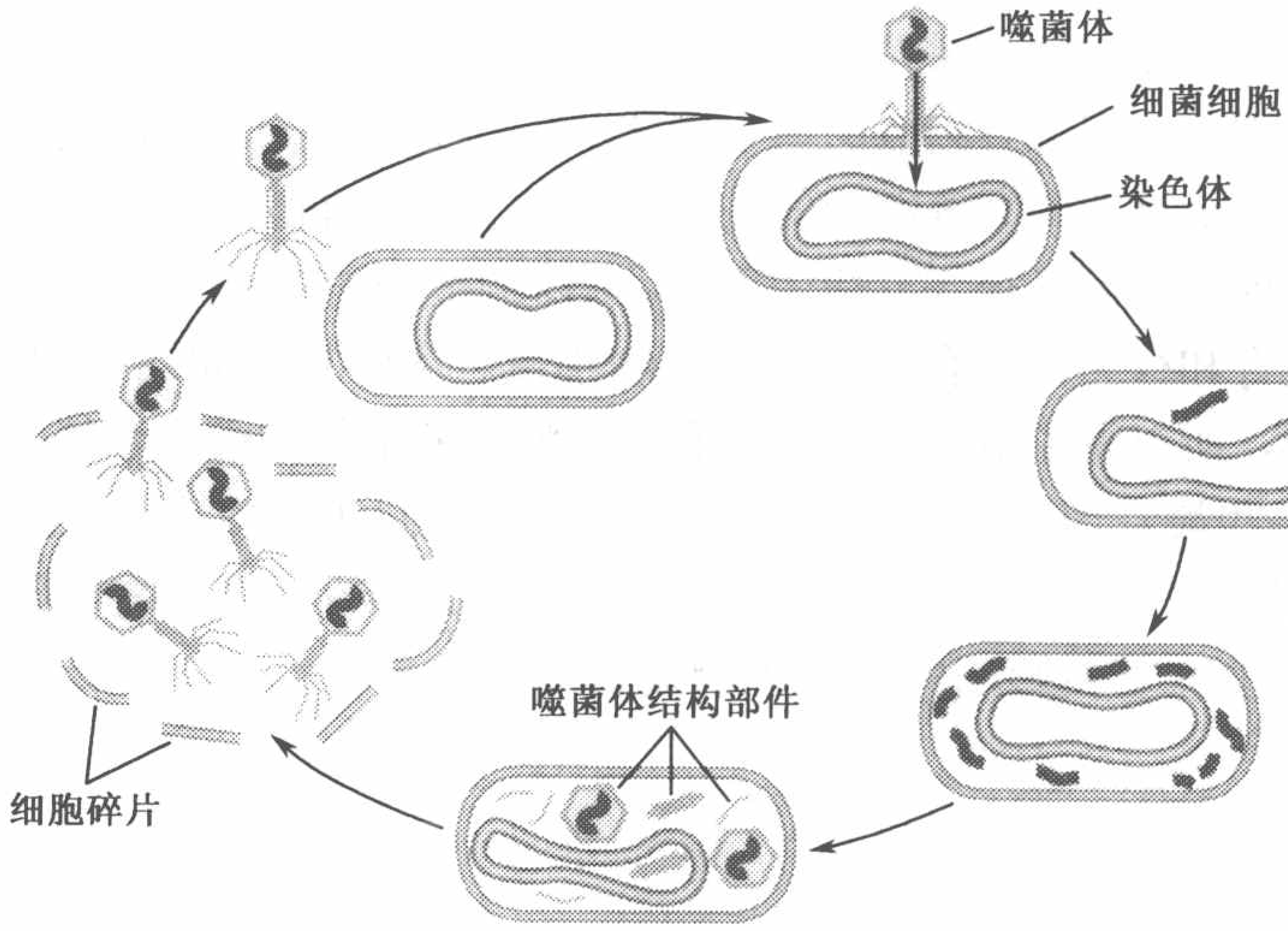


图 22-1 噬菌体的裂解生长周期。噬菌体颗粒附着在合适的细菌宿主细胞（带有合适的受体）的外表面，并将其基因组注入，通常是 DNA 分子。噬菌体 DNA 在细菌中复制，基因表达产生许多新的噬菌体。一旦子代噬菌体装配为成熟的颗粒，细菌细胞裂解，子代噬菌体释放，感染其他的宿主细胞。

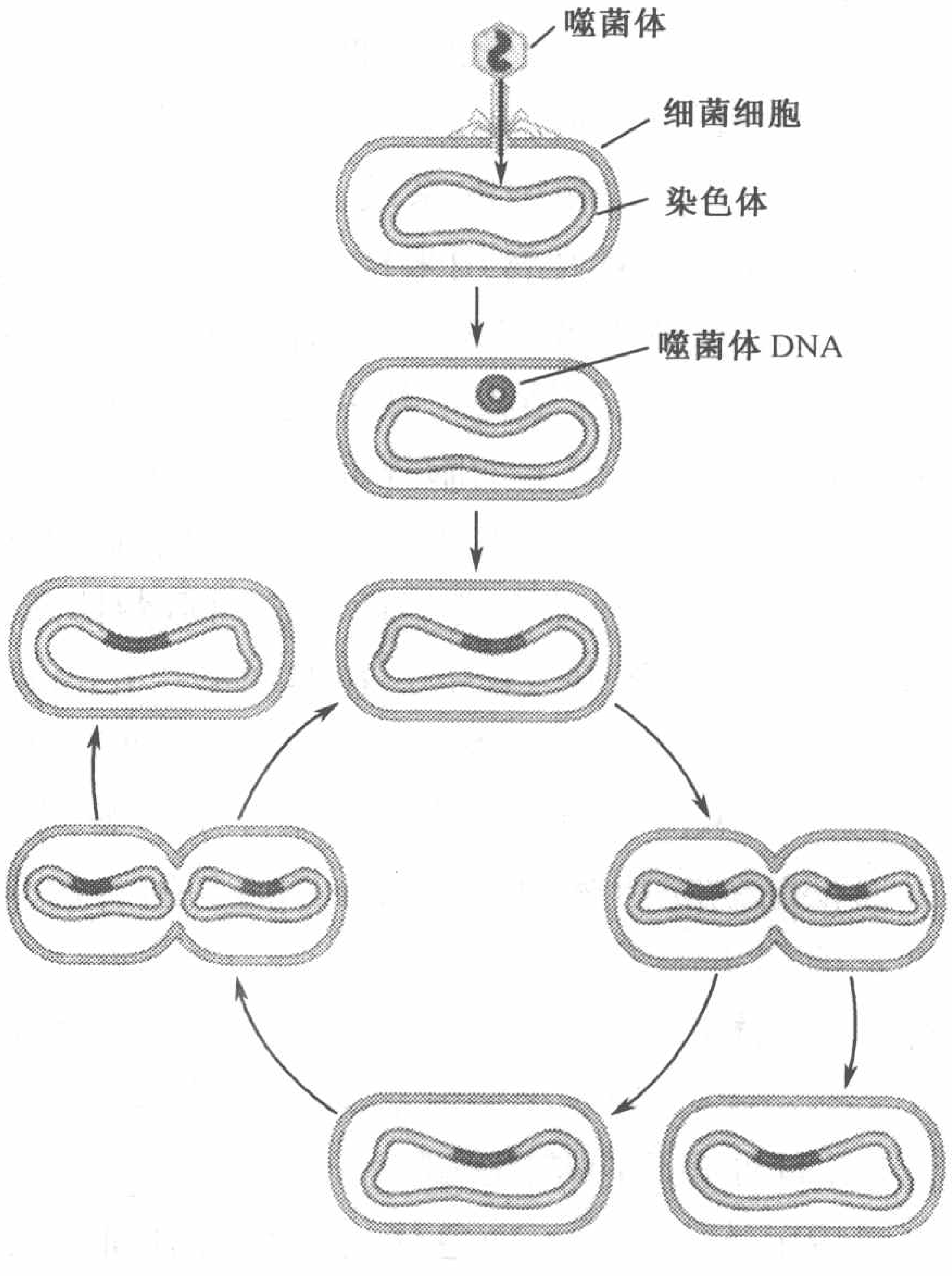


图 22-2 噬菌体的溶原周期。最初的感染步骤和裂解周期的情况一样（图 22-1）。但 DNA 一旦进入细胞，就整合到细菌染色体上，作为细菌染色体的一部分而被动复制。而且，编码外壳蛋白的基因处于关闭状态。整合的噬菌体叫原噬菌体。溶原性细菌可以稳定地保持若干代，在合适的环境下，也可以有效地转为裂解周期。详见第 16 章。

噬菌体生长的检测

要想使噬菌体成为有用的实验系统，就需要建立噬菌体扩增和定量的方法。扩增可以产生高滴度噬菌体材料——用于进行实验或提取 DNA。一般情况下，噬菌体可以用液体培养基在合适的细菌宿主里生长而得以扩增。例如，一瓶生长旺盛的细菌细胞被噬菌体感染。生长一段合适的时间后，细胞裂解，留下透明的含有噬菌体颗粒的液体悬浮液。

噬菌斑检测（图 22-3）用来定量噬菌体颗粒数目。它的步骤是：混合噬菌体和细菌细胞，噬菌体吸附于细菌细胞上并将 DNA 注入；然后稀释噬菌体和细菌混合物，稀释液加入到含有更多（没有被感染）的细菌细胞的“软琼脂”里，倒在培养皿中硬琼脂培养基上。软琼脂形成一层果冻状的薄层，里面混有细菌。有些是被感染的，但大多数不是。然后将平板培养几个小时，使细菌生长和噬菌体感染得以进行。

在接下来的培养过程中，软琼脂里的每一个被感染的细胞（来自最初的混合体）将会裂解。琼脂的硬度使子代噬菌体扩散，但扩散不远，所以它们只感染生长在其周围的细菌细胞。随后，这些细胞裂解释放更多的子代，再次感染相邻的细胞，如此循环往复，多轮感染的结果是噬菌斑（plaque）形成。在高密度生长的未被感染的细菌细胞组成的不透明的菌苔背景下噬菌斑呈透明的圆斑。这是因为未被感染的细菌细胞在软琼脂里生长成为高密度群体，而那些位于每个最初感染的细菌周边的细菌细胞被杀死，留下一个透明的斑块。知道了一块特定平板的噬菌斑数目，以及在铺平板之前最初的噬菌体的稀释程度，计算噬菌体在原液中的数目就轻而易举了。

一步生长曲线

这个经典的实验揭示了一个典型的裂解型噬菌体的生命周期，为许多检验生命周期的后续实验铺平了道路。这个过程的基本特征是对一群细菌同步感染并避免子代的再次感染。这样就可以跟踪单轮感染的过程（图 22-4）。

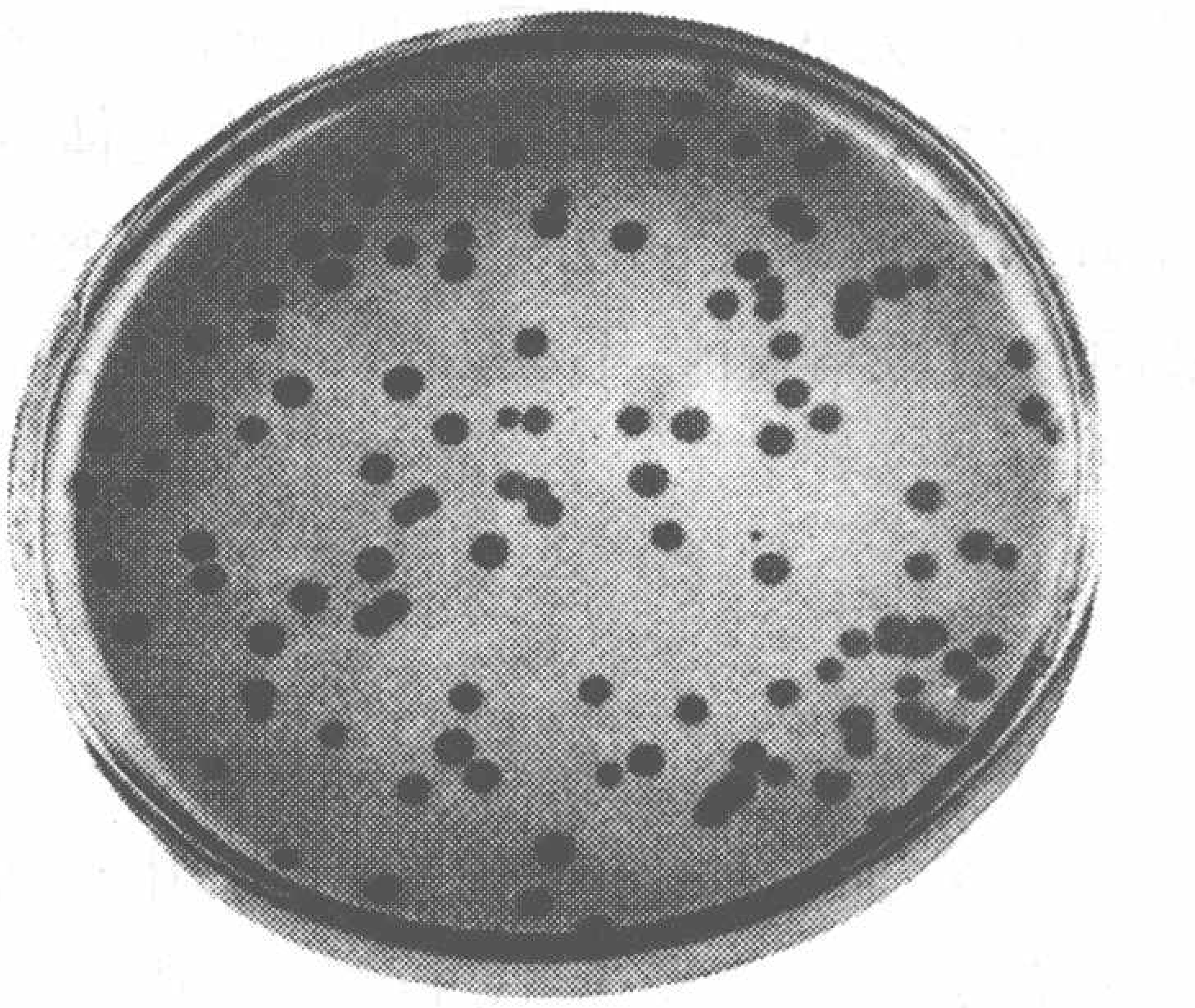


图 22-3 噬菌体感染细菌菌苔形成噬菌斑。所示为裂解型 T 噬菌体产生的噬菌斑。（经许可，影印自 Stent G. S. 1963. *Molecular biology of bacterial viruses*, Fig. 1. © W. H. Freeman.）

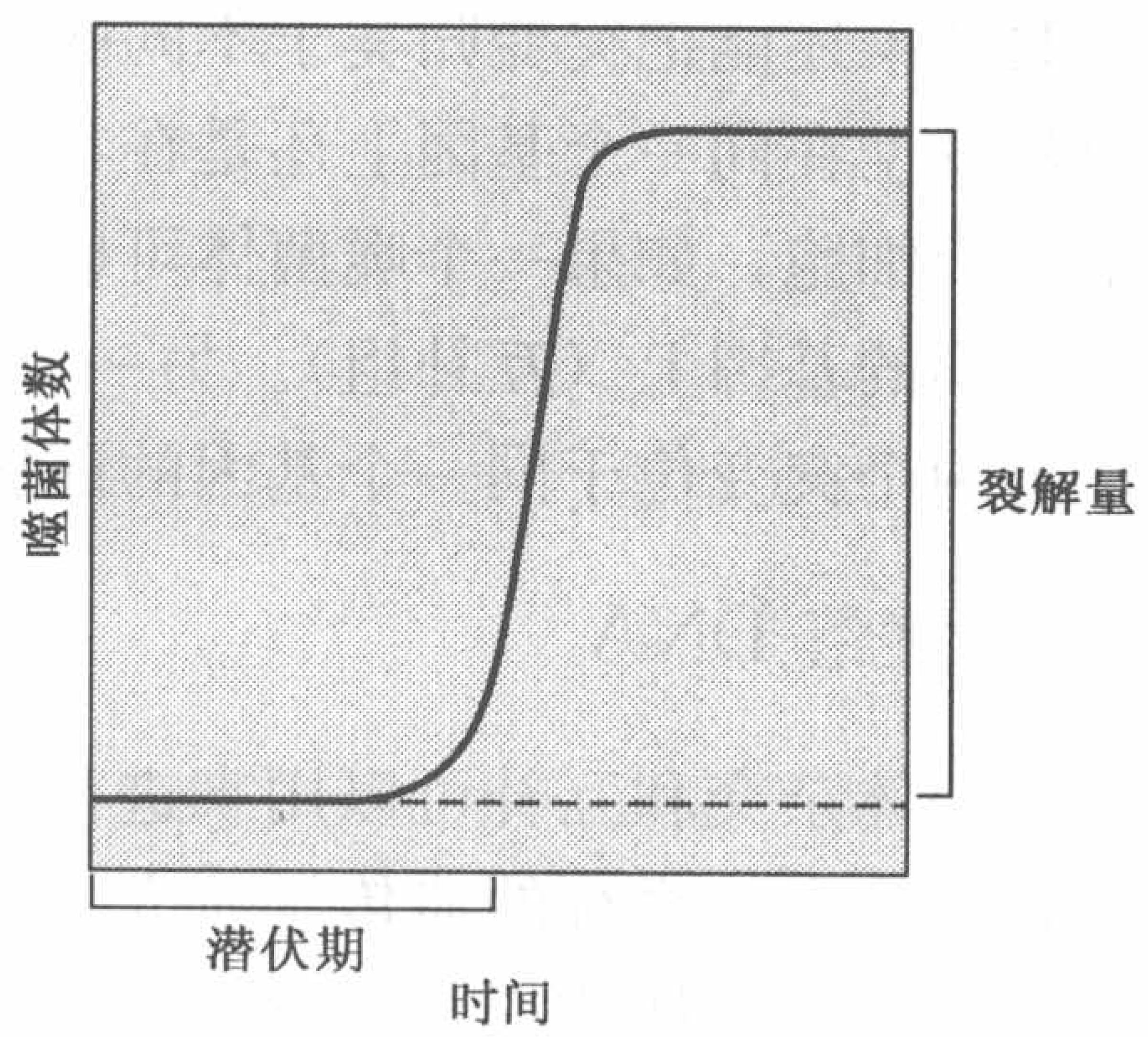


图 22-4 一步生长曲线。如文中所描述，一步生长曲线显示一个噬菌体进行一轮裂解生长所用的时间，以及每个被感染细胞产生的子代噬菌体数目。它们分别是潜伏期和裂解量。

噬菌体和细菌细胞混合 10min，这段时间足以使噬菌体吸附到细菌细胞上，但对进一步感染来说则太短。然后把这个混合物稀释（用新鲜的培养液）10 000 倍。这个稀释度保证只有那些最初就结合噬菌体的细胞才能形成感染群体，同时，保证产生的子代噬菌体不能感染宿主细胞。

稀释的感染细胞群随后进行培养使感染过程继续下去。每隔一段时间，从混合物中取出样品，用噬菌斑检测法计数噬菌体的数目。起初，噬菌体数目很少（仅仅由稀释前的初始感染时未能感染细胞的噬菌体组成）。

一旦有足够的时间使被感染的细胞裂解并释放子代，就会检测到噬菌体数目的剧增（对 T4 噬菌体来说这个过程需要 30min）。从感染到释放子代噬菌体的时间间隔叫做潜伏期（latent period），释放的噬菌体数目叫做裂解量（burst size）。

噬菌体杂交和互补实验

对一个群体中噬菌体数目的计数可以使研究者了解一个特定的噬菌体衍生物是否能够在一个特定的宿主细菌中生长（以及生长效率，如裂解量）。而且，平板检测可以分辨不同种类的噬菌体衍生物，因为它们产生的噬菌斑形态不同。宿主范围和噬菌斑形态的差别通常是在其他方面都相同的噬菌体在遗传上有所不同的结果。在分子生物学的早期，上述差异提供了一种遗传标记，使研究者们能够探询遗传信息是如何编码和行使功能的。

混合感染——1 个细胞同时被 2 个噬菌体颗粒感染，从两个方面使遗传分析成为可能。第一，使噬菌体杂交得以进行。如果同种噬菌体的两个不同的突变体（带有同源染色体）共同感染一个细胞，重组即遗传交换就可以在基因组之间发生。遗传交换的频率可以用来在基因组上排序基因。重组频率高表明突变的位置相对较远，而重组率低则表明突变的位置较近。实验中可以处理大量噬菌体，只要有筛选或选择这种小概率事件的方法，即使概率非常小的事件（2 个位置非常相近的突变之间的重组）也可以发生。第二，共感染还能把突变归类于不同的功能互补组，也就是说，人们可以确定两个或多个突变是发生在同一个基因上还是在不同的基因上。这样，两个不同的噬菌体突变体共同感染一个细胞，如果一个噬菌体可以提供另一个噬菌体所缺乏的功能，那么两个突变即位于不同的基因上（互补组）。另一方面，如果两个突变噬菌体不相互互补，那么就可以作为两个突变位于同一个基因的证据。

转导和重组 DNA

噬菌体杂交和互补可以用来进行噬菌体本身的遗传学实验，此类相同的载体和技术也可以用于其他系统的遗传学研究。起初，仅仅是观察到一些噬菌体感染过程中意外获得的细菌基因（如我们在下面描述的）。然而，随着 20 世纪 70 年代 DNA 重组技术的出现，这些研究扩展到来自各种生物的 DNA。

在感染过程中，噬菌体可能意外地（和错误地）获得一段细菌的 DNA，通常，噬菌体获得一段宿主细菌 DNA 是在原噬菌体溶原诱导过程中从细菌染色体剪切出来的。该过程涉及位点特异的重组事件（第 11 章），只要重组发生一点点位置错误，就会有噬

菌体 DNA 的丢失和细菌 DNA 的获得。只要这个交换没有去掉噬菌体基因组中增殖所需的部分，形成的重组噬菌体仍然可以生长，并可以把获得的细菌 DNA 从一个细菌宿主转移到另一个细菌宿主，这个过程被称为特异性转导（specialized transduction）。包含在特异性转导噬菌体中的细菌 DNA 也同样可以进行像噬菌体一样的遗传分析。

因为 λ 噬菌体具有促进特异性转导的能力，所以它很自然的被选为原始的克隆载体之一（第 21 章）。去除特定限制性内切核酸酶的多余位点，仅在对裂解生长无关紧要的区域留下一个（插入型载体）或两个（置换型载体）位点，就可以使 λ 噬菌体能够接受任意来源 DNA 的插入（体外）。如此进行 DNA 的扩增和分析，会比在原来的生物体内容易得多。去除 λ 噬菌体中限制性内切核酸酶位点是反复选择对某种限制性系统具有越来越高效率的菌株而实现的。用这种方法增加对限制性内切核酸酶的抗性，然后，在体外定位丢失和保留的位点，鉴定出所需的噬菌体衍生物。

人们先后开发了多种 λ 载体，其差别在于使用的限制位点以及如何鉴定重组噬菌体。下面介绍一种选择系统的作用机制：在一个 λ 的衍生物中，编码阻抑物的 *cI* 基因里（第 16 章）留下一个唯一的某种限制性切点。在亲代载体中，这个基因是完整的，噬菌体可以根据自己的选择形成溶原性细菌，于是噬菌体形成的是浑浊的噬菌斑。然而，当一段 DNA 插入在这个位点，形成的重组噬菌体其 *cI* 基因被打断，不能形成溶原性细菌，所以形成透明的噬菌斑。

噬菌斑形态的改变使从非重组噬菌体中辨别重组子非常简便，进一步讲，如果使用的细菌菌株是 *hfl* 菌株（框 16-5），这个途径就是选择（而不是筛选）。用该菌株，任何溶原性噬菌体总是溶原的，而重组的噬菌体会形成噬菌斑。

细菌

细菌，如大肠杆菌和芽孢杆菌，作为实验系统的迷人之处在于它们是相对简单的细胞，并且可以相对容易地培养和操作。细菌是单细胞生物，所有 DNA、RNA 和蛋白质合成的机器都包含在同一细胞器中（细菌没有细胞核）。

细菌通常只有一条染色体——比高等生物的基因组小得多。此外，细菌的生命周期很短（有些甚至只有 20min），从单个细胞可以很容易获得一个遗传上同源的细胞群体（克隆）。最后，细菌遗传学研究很方便，一方面，它们是单倍体（意味着即使是隐性突变，也能够表现出突变的表型）；另一方面，细菌之间可以方便地进行遗传物质的交换。

分子生物学起源于细菌和噬菌体模式系统的实验。至 1943 年 Salvadore Luria 和 Max Delrück 进行著名的波动分析实验以前，细菌研究（细菌学）还停留在传统遗传学领域之外。利用统计学的方法，Luria 和 Max Delrück 证明了细菌可以发生某种变化，使其对某一特定噬菌体的感染具抗性。更重要的是，他们证明了这种变化是自发产生的，而不是对噬菌体的应答（适应）。所以，像其他生物一样，细菌也可以遗传性状（如对噬菌体的敏感性或抗性），并且偶尔这种遗传性可以自发产生变化（突变）而成为另一种遗传状态。Luria 和 Max Delrück 的实验表明，同其他生物一样，细菌表现出的

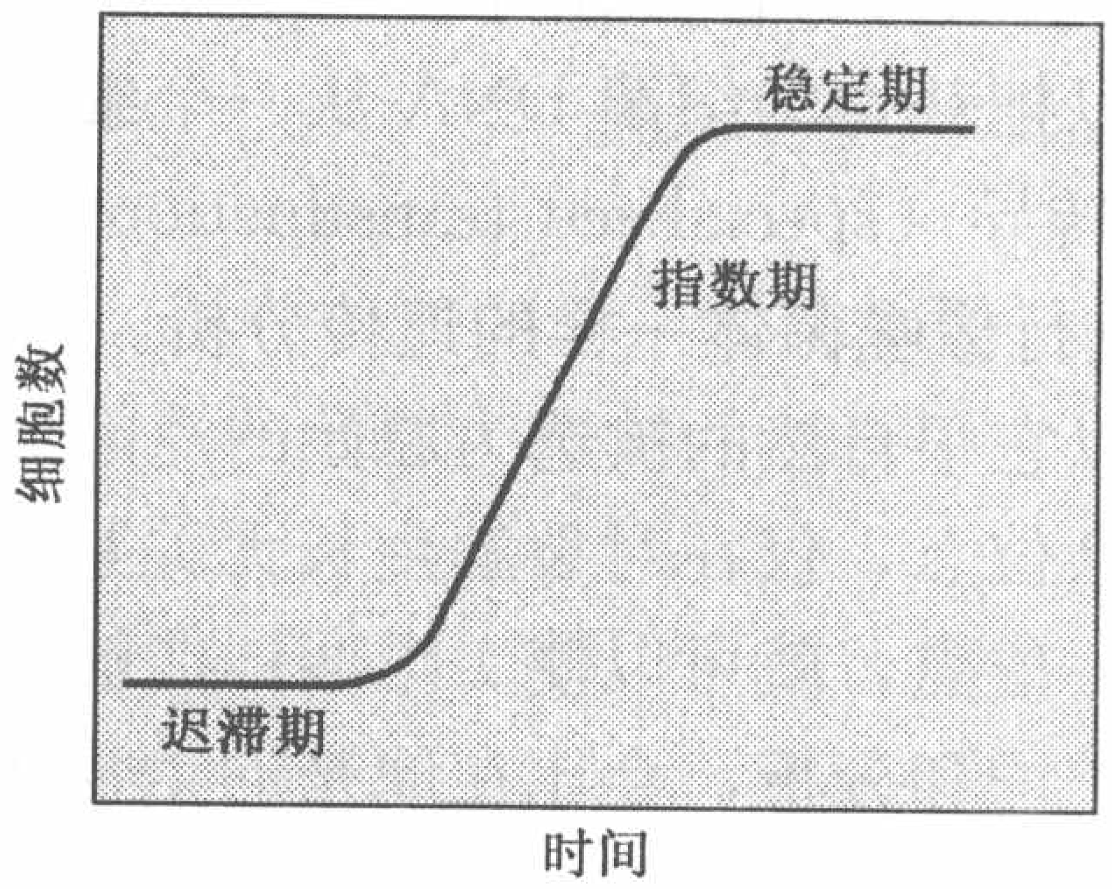


图 22-5 细菌生长曲线。 如文中所描述，细菌（如大肠杆菌）在不过度密集并且在氧气充足的丰富培养基里繁殖时生长速度很快。这个生长期叫指数生长期，因为细胞以指数级复制。一旦细胞的数目变得太高，培养物变浓，生长就变慢，进入所谓的稳定期。取出处于稳定期的细胞，在新鲜培养基中将它们稀释，经过一个迟滞期后，细胞将再次进入指数生长期。图示为不同时期细胞数目增长的速度。

特性是由遗传因素决定的。因为它们结构简单，细菌是阐明遗传物质本质和孟德尔（Gregor Mendel）性状决定因子（基因）的理想实验系统。

细菌生长的检测

细菌可以在液体或固体（琼脂）培养基中生长。细菌细胞的大小（大约 2μm 长）足以使它散射光线，因此通过检测细菌培养物中光密度的增加就可以监控细菌的生长情况。以恒定的繁殖世代分裂的活跃生长的细菌其数量以指数级增长，此时它们被称为处于生长的指数期（exponential phase of growth），当细菌数量增长到一定程度后，生长速度减缓，细菌生长进入稳定期（stationary phase）（图 22-5）。

细菌的数目可以通过把培养物稀释然后铺在培养皿中的固体（琼脂）培养基上进行测定。在相对较短的时间内，单个细胞就能长成由上百万个细胞组成的微型菌落，只要知道平板上的菌落数和培养物的稀释倍数就可以计算出原始培养物中的细菌浓度。

细菌通过性结合、噬菌体介导的转导和 DNA 介导的转化交换 DNA

作为分子生物学的模式系统，细菌的主要优势是具备易于进行遗传交换的系统。遗传交换使定位突变、构建含多种突变的菌株、构建用来辨别显性突变和隐性突变及进行顺反式分析的部分双倍体的菌株成为可能。

大部分细菌含有自主复制的 DNA 元件，称为质粒（plasmid）（图 22-6）。有些质粒，如大肠杆菌的育性质粒（称为 F 因子，F-factor）具备把它们自身从一个细胞转移到另一个细胞的能力。含有 F 因子的细胞（F^+）可以把质粒转移到 F^- 细胞里。F 因子介导的接合是一个复制的过程。所以 F^+ 细胞转移一个拷贝的 F 因子给 F^- 细胞，自身还有一个拷贝，这样接合的结果就是产生 2 个 F^+ 细胞。有时，F 因子整合到染色体中，就会引起宿主染色体通过接合向 F^- 细胞转移。含有整合的 F 因子的菌株叫做 Hfr（高频重组）菌株（Hfr strain），这种材料对进行遗传交换研究非常有用。

在某一具体的事件中，决定宿主染色体转移的精确位置取决于两个条件。第一，不同的 Hfr 菌株中 F 质粒整合在宿主染色体的不同位置。宿主染色体到受体细胞的转移是沿着线形发生的，从接近整合有 F 质粒的那部分染色体区域开始。所以，质粒整合的位置决定了首先被转移的染色体位置。第二，在交配中断之前完成整个染色体的转移是少见的。因而，远离转移起始点的基因转移频率低，在一个特定的交配中，位置较远的基因可能从来不会被转移。需要注意的是，如果一个完整拷贝的整合 F 因子发生转

移的话，发生在最后。

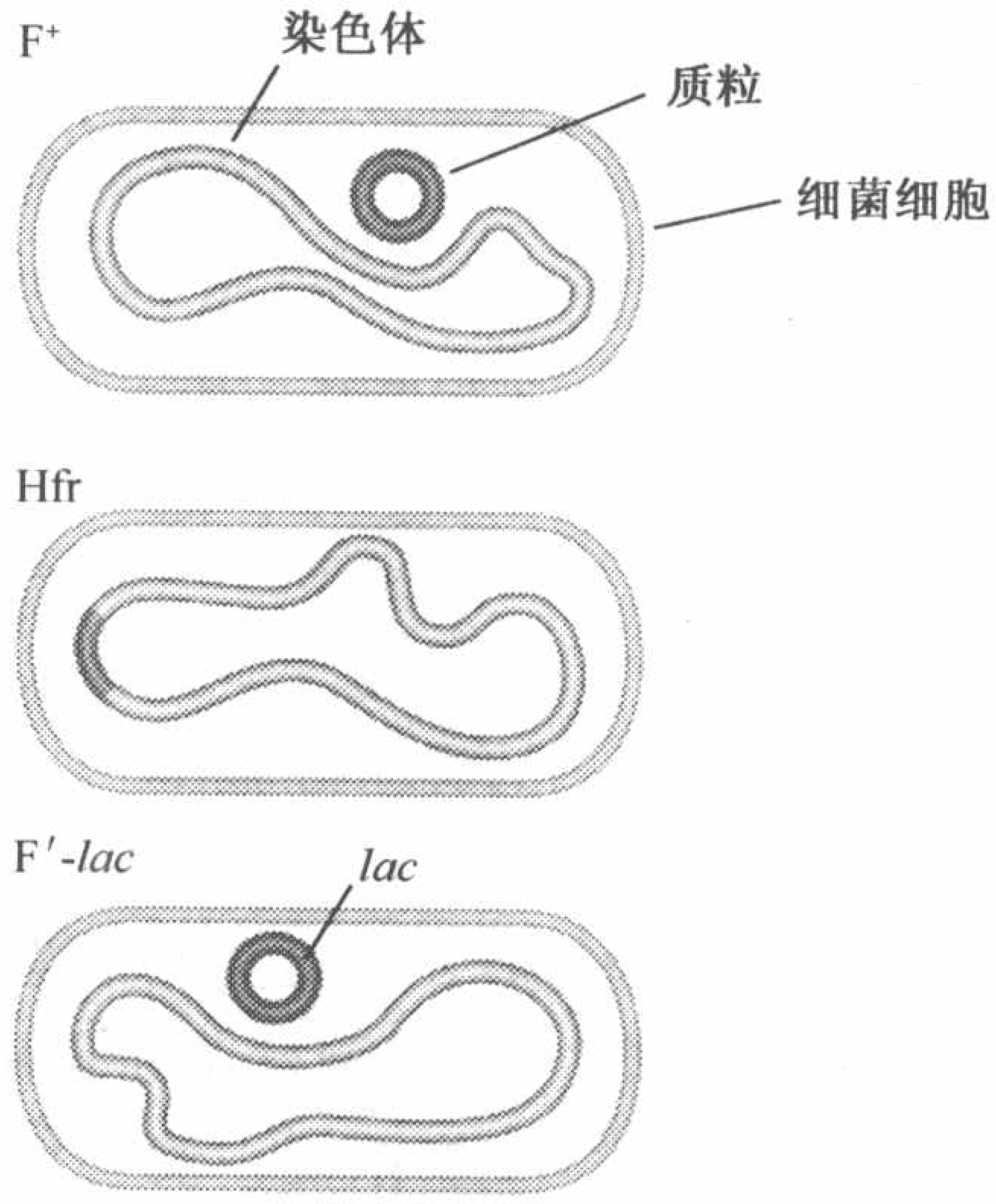


图 22-6 带有 F 质粒的 3 种细胞形式。F^+细胞带有一个单拷贝的 F 质粒，F 质粒就像一个微小的染色体独立进行复制。在 Hfr 菌株中，F 质粒整合在细菌染色体里作为染色体的一部分而复制。在 F′菌株里，已经整合到宿主染色体中的 F 质粒重新游离出来，同时携带一小段相邻的宿主 DNA。所有 3 种细胞类型都可以转移到一个 F^- 受体细胞。如果供体细胞是 F^+ 菌株，它仅复制和转移 F 质粒；如果是 F′菌株，它复制和转移 F 质粒以及一小段宿主 DNA；如果是 Hfr 菌株，根据整合的位置和交配的时间不同，它复制和转移不同大小和位置的宿主染色体。一旦进入受体，来自宿主的染色体 DNA 就可以与受体细胞的基因组重组，从而产生了遗传交换。

第三种也是非常重要的 F 因子形式是 F′质粒，F′是含有一小段染色体 DNA 的育性质粒，这段 DNA 会与质粒一起从一个细胞向另一个细胞高频率转移。例如，历史上重要的 F′-*lac*，是一个带有乳糖操纵子的 F 因子。F′因子可以产生含有部分双倍体的菌株，这个菌株的特定染色体区域有两份拷贝。这正是 Jacob 和 Monod 用来制备进行顺反分析的部分双倍体菌株的方法，这些菌株中突变是发生在乳糖操纵子抑制基因以及与抑制物结合的启动子位点上（框 16-3）。

F 因子只能与其他大肠杆菌菌株接合，然而，有些其他的接合性质粒没有选择性，可以将 DNA 转移到多种不相关的菌株中，甚至酵母里。这些没有选择性的接合性质粒提供了一个将 DNA 包括通过重组 DNA 技术得到的 DNA，引入到自身没有遗传交换系统的细菌菌株的简便办法。

遗传交换的另一个强有力的工具是噬菌体介导的转导（图 22-7），普遍性转导（generalized transduction）是由那些在病毒成熟时偶然包裹一段染色体 DNA 而非病毒 DNA 的噬菌体所介导。当这样的噬菌体颗粒感染下一个细胞时，从以前宿主那里获得的染色体 DNA 片段将取代感染性病毒 DNA。注入的染色体 DNA 可以和被感染的宿主细胞的染色体重组，导致遗传信息从一个细胞到另一个细胞的永久性转移，这种转导称作普遍性转导，因为宿主染色体 DNA 的任何片段都可以从一个细胞转移到另一个细

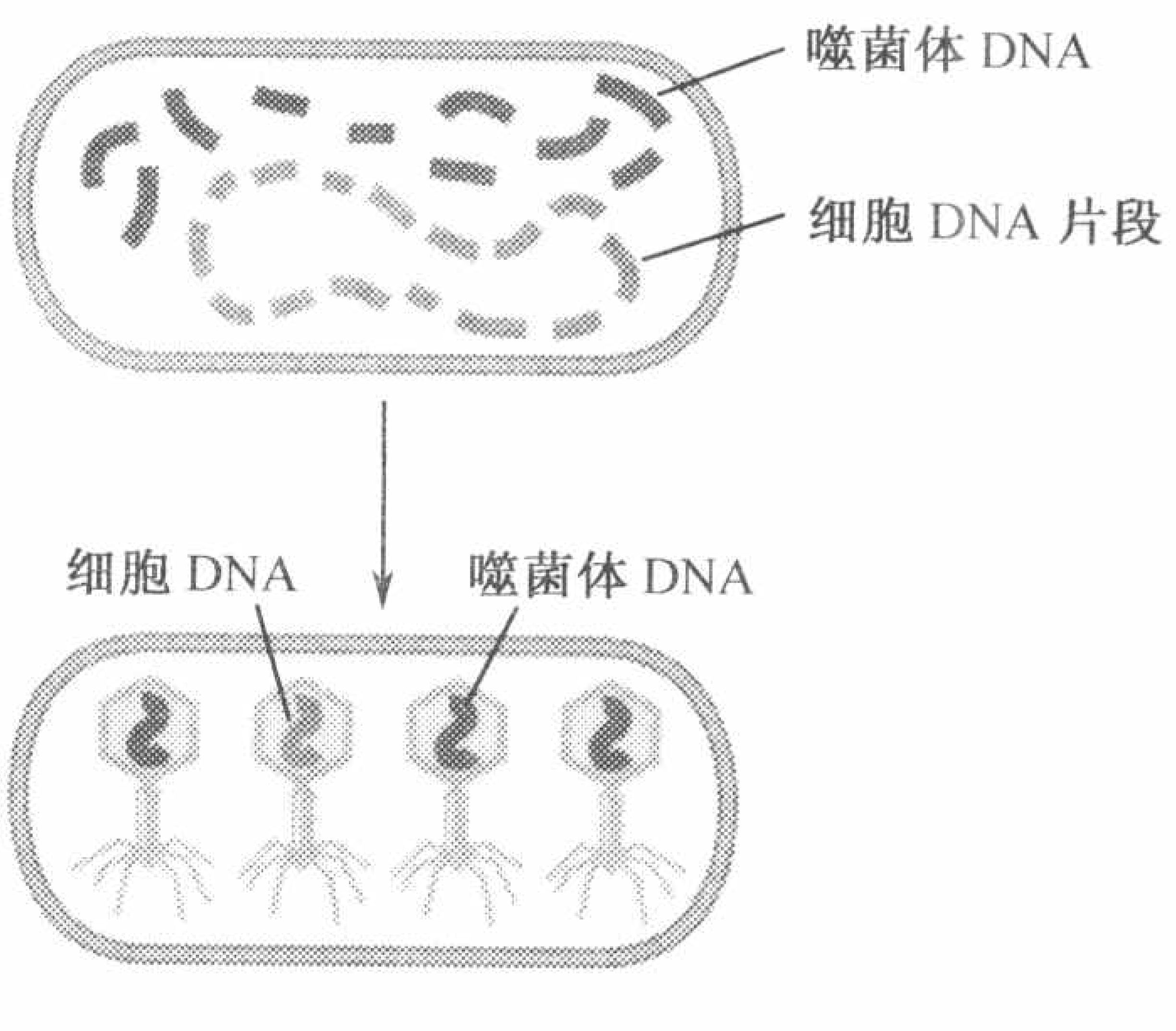


图 22-7 噬菌体介导的普遍性转导。如文中所述，在有些噬菌体感染时，宿主染色体被打成片段，有些 DNA 片段可以取代复制的噬菌体 DNA 被包装在噬菌体颗粒里。与噬菌体基因组类似，宿主 DNA 被转移到另一个细胞中。一旦进入新的宿主，DNA 就可以与染色体重组，促进遗传交换。

胞。根据病毒的大小，有些普遍性转导噬菌体只能转导几千个碱基的染色体 DNA，而有些则可以转导大于 100kb 的 DNA。

正如已经提到过的，另一种噬菌体介导的转导叫做特异性转导（specialized transduction）。这个过程涉及溶原性噬菌体，如某段噬菌体 DNA 被一段染色体 DNA 所取代的 λ 噬菌体。一旦感染细菌，这样一个特异性转导噬菌体就可以将这段细菌 DNA 转移到新的细菌细胞中。

最后，我们来看 DNA 介导的转化，对此第 21 章已有所描述。某些重要的实验细菌（如芽孢杆菌而不是大肠杆菌）有天然的遗传交换系统，使它们可以通过重组吸收并整合线状的、裸露的 DNA（它们的同胞释放的或从它们的同胞那儿获得的）到它们自己的染色体里。通常这些细胞必须处于“遗传感受态”才能从环境中吸收和整合 DNA。遗传感受态特别重要，可以先用重组 DNA 技术来克隆的染色体 DNA 片段，然后使它被感受态细胞摄取并整合到受体细胞的染色体中。

细菌质粒可以用作克隆载体

就我们所知，细菌通常含有环状的质粒 DNA，它们可以自主复制，这样的质粒可以方便地作为细菌 DNA 和外源 DNA 的载体。实际上，最初的（和成功的）克隆重组 DNA 的实验就使用了一个大肠杆菌的质粒（pSC101），它含有一个唯一的 *Eco*RI 限制位点，DNA 可以插入这个位点而不影响质粒复制的能力（第 21 章）。

转座子可用来产生插入突变和基因与操纵子的融合

在第 11 章中讨论过，转座子（transposon）不仅凭其自身的能力而成为一个重要的遗传元件，而且在细菌分子遗传操作中也非常有用。例如，以低序列特异性（即高度随机性）整合在染色体上的转座子，如 Tn5 和 Mu，可以用来产生全基因组水平的插入突变体文库（图 22-8）。

这种突变与传统的化学诱变产生的突变相比有两个重要的优势，一是转座子插入到基因中（当想要的时候）很可能导致基因的完全失活（功能丧失的突变），而不像诱变剂只能产生简单的碱基改变；第二个优势是，插入 DNA 不但可以失活基因，而且使分离和克隆相应基因方便易行。更简单的办法是，可以用合适的 DNA 引物，通过分析有转座子插入的染色体 DNA 序列而确定失活基因。

转座子还可以用来在全基因组水平上产生基因和操纵子的融合。经过改造，构建了带有报道基因的转座子，如没有启动子的 *lacZ*（如 Tn5*lac*）转座子。当这种转座子

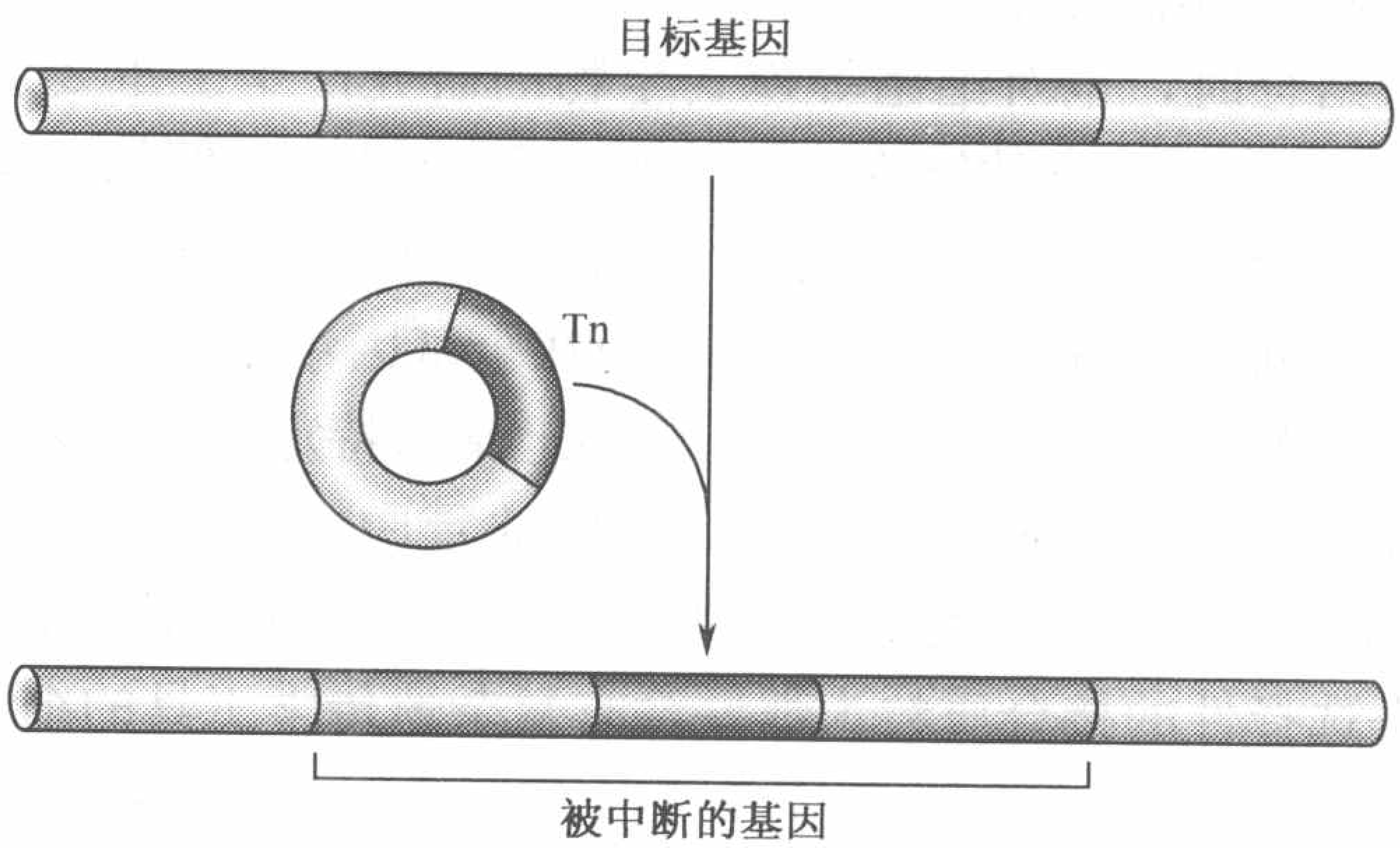


图 22-8　转座子产生的插入突变。转座子被质粒携带进入细胞，随后从载体上转移到宿主染色体上。因为在细菌染色体中编码区域（基因）高度密集，转座子插入基因的频率很高。转座子中携带的标记（如抗生素抗性）使带有转座子的细胞可以被分离。知道转座子两端的序列以及它所插入的基因组的序列，就可以很方便地鉴定转座子所处的位置。

（以合适的方向）插入到染色体里，报道基因的转录就会受到被打断的目标基因的控制。这样的融合称为操纵子或转录融合（图 22-9）。

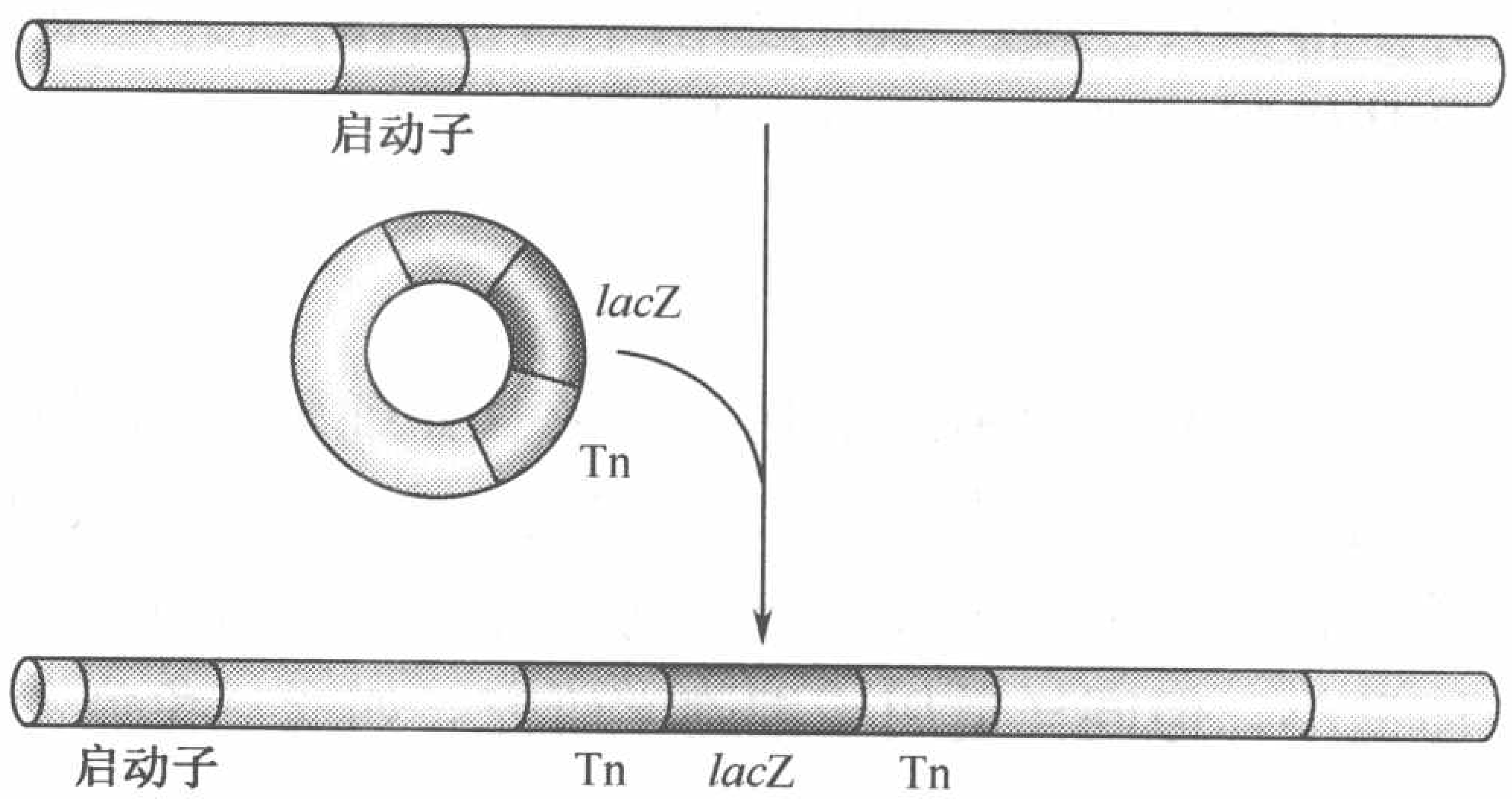


图 22-9　转座子产生的 *lacZ* 融合。将前图介绍的转座子诱变法稍加修改，就可以用来将报道基因（如 *lacZ*）插入到基因组的任何区域。这样只要通过测量菌株中 *lacZ* 的表达水平就可以检测目标基因（被转座子-*lacZ* 融合插入的基因）的表达。

其他用于产生融合的转座子已制备出来了，它们含有既无启动子又无翻译起始序列的报道基因。在这些转座子中，报道基因需要置于目标基因的转录控制下并且在能够被正确翻译的目标基因可读框内才能表达。报道基因同时在转录和翻译水平上参与目标基因的融合称为基因融合。

重组 DNA 技术、全基因组测序和转录谱绘制加强了细菌的分子生物学研究

重组 DNA 技术，如 DNA 克隆、全基因组序列的完成以及在全基因组水平研究基

因转录的方法不但使高等细胞的分子生物学研究产生了革命性的突破，尤其是在与传统的细菌遗传学紧密结合时，也对细菌等模式生物的研究产生了重要影响。例如，重组DNA 技术使制备用于基因融合的转座子衍生物变得容易；又例如，结合遗传感受态与重组方法可以产生精确的突变和基因融合，大大增加分子遗传操作的种类和数目。包含细菌中所有基因微阵列（Microarray）的利用使在全基因组水平上研究基因表达成为可能。结合前面描述的各种技术，可以快速、简便地阐明特定条件下表达的所有基因的功能。已经在酵母和其他真核系统中发挥巨大作用的蛋白质间相互作用的快速鉴定方法（如双杂交分析，框 17-1），也是阐明细菌蛋白质之间相互作用网络的强有力工具。全基因组序列和没有选择性的接合质粒为那些没有复杂遗传学工具的细菌提供了进行分子遗传学操作的机会。

有了更好的传统和分子遗传学的工具，简单细胞的生化分析就会更加有效

从分子生物学的最初建立起，细菌就处于 DNA 复制、信息传递、基因调节机制等各种生物化学研究的中心。这有几个理由：第一，可以在确定的和均一的生理条件下培养大量的细菌细胞；第二，传统和分子遗传学的技术使纯化或大量生产经过改造的蛋白质复合物成为可能，从而可以获得大量的单个蛋白质；第三，也是非常重要的，就像在本书中反复提到的，细菌的 DNA 复制、基因转录、蛋白质合成等机器比高等细胞简单得多（所包含的成分少得多）。所以，阐明细菌基本机制的研究进展迅速，只需要分离少数的蛋白质，而且其机制一般比高等细胞更简单。

细胞学分析同样适用于细菌

尽管细菌形态简单，没有膜包裹的细胞内分区（如细胞核和线粒体），但细菌并不是像几十年来人们认为的那样，只是简单的一包酶。实际上，正如我们现在已经知道的，蛋白质和蛋白质复合物在细胞中都有特定的位置，连细菌中的染色体都是高度有组织的。尽管细菌的体型小，但同样适用于细胞学研究，如通过免疫荧光显微镜用特定抗体定位固定细胞内的蛋白质，用荧光显微镜与绿色荧光蛋白（green fluorescent protein）可以定位活体细胞中的蛋白质，还有用荧光原位杂交（FISH）可以定位细胞内染色体区域和质粒。这些方法的应用对我们深刻地理解本书所讨论的若干分子过程很有帮助。例如，我们现在知道，细菌细胞的复制机器是相对固定的并且定位在细胞中央（第 8 章）。这个发现告诉我们在复制过程中，DNA 模板线状地通过一个相对固定的复制“工厂”，而不是传统观点认为的 DNA 聚合酶像火车在铁轨上那样沿着模板行进。又如，应用细胞学方法使我们清楚了（同样与传统的观点相反）复制过程中染色体的新复制的两个起始区域朝细胞的两极移动。细胞学方法是细菌分子生物学研究的重要手段之一。

关于基因的基本知识绝大部分来自噬菌体和细菌

分子生物学起源主要归功于用细菌和噬菌体等模式系统进行的实验。如同我们在第

2 章所看到的，对于肺炎球菌的开创性研究使人们发现遗传物质就是 DNA。此后，就像我们在本书中所看到的，大肠杆菌及其噬菌体的实验一直在引导学科的发展。例如，Hershey 和 Chase 的实验使人们相信噬菌体的遗传物质是 DNA；Meselson 和 Stahl 的实验证明在大肠杆菌中 DNA 进行半保留复制；Crick 和 Brenner 的噬菌体交配实验（第 15 章）揭示了遗传密码由三联体密码子构成；而 Yanofsky 用大肠杆菌进行的出色的遗传研究证明了遗传的共线性；也不能忘记 Jacob 和 Monod 的工作（框 16-3），是他们发现了基因调控的基本机制；此外还有无数其他例子。在这些例子里，通过研究这些最简单的系统，生命的基本过程得以揭示。

一个重要的例子来自 Seymor Benzer 的出色工作，他详细研究了 T4 噬菌体中的一个叫 *rII* 的遗传位点。野生型 T4 可以在大肠杆菌 B 或 K 菌株里生长，但 *rII* 突变体仅在 B 菌株中生长。这样就有可能检测到发生频率小于 0.01%的野生型噬菌体（比如说来自两个不同 *rII* 突变体之间的重组）。也就是说，当把 T4 噬菌体铺在 K 菌株形成的菌苔上时，可以在 10 000 个 *rII* 突变噬菌体中检测到一个野生型，因为只有这种频率极低的重组子才能形成噬菌斑。

利用 *rII* 突变体这一难以置信的特性，Seymour Benzer 在成对的 *rII* 突变体之间进行了重组实验，并因此能够以高分辨率（接近或达到核苷酸碱基对的分辨率）绘制这些突变的顺序。他还设计了一个“互补”实验（上述）表明 *rII* 位点是由两个相邻的基因组成的。Benzer 引入了顺反子（cistron）这一术语来描述基因（基于顺和反这两个词）。此外，值得一提的是，正是这个工作使 Crick 和 Brenner 能够利用这一位点开展遗传密码的研究。

酿酒酵母

单细胞真核生物作为模式实验系统有许多优势：比起其他真核生物，它们的基因组较小（第 7 章），基因数目也比较少；与大肠杆菌相似，它们可以在实验室里快速繁殖（理想条件下，每次细胞分裂需要大约 90min），从单个细胞繁殖成克隆群体。尽管酵母细胞比较简单，但它们具备所有真核生物细胞的主要特征，如含有一个独立的细胞核、多条线性染色体包装成染色质、细胞质包含了全部的细胞器（如线粒体）和细胞骨架结构（如肌动蛋白纤维）。

研究得最清楚的单细胞真核生物是酿酒酵母 *S. cerevisiae*，因为可以用作发酵剂，所以通常称为酿酒酵母。100 多年来人们对酿酒酵母进行了大量的研究。19 世纪 60 年代，在实验中，路易斯 · 巴斯德鉴定出这种酵母是发酵的催化剂（糖被认为是自发地分解成乙醇和二氧化碳）。这些研究最终导致了第一个酶的发现以及生物化学作为实验体系的建立。20 世纪 30 年代，开始了酿酒酵母的遗传学研究，由此鉴定了许多酵母基因。所以，像大肠杆菌一样，酿酒酵母使研究者能用遗传和生化的方法来解决生物学基本问题。

单倍体和双倍体细胞的存在促进了酿酒酵母的遗传分析

酿酒酵母细胞可以以单倍体状态（每条染色体有一个拷贝）或双倍体状态（每条染色体有两个拷贝）生长（图 22-10）。单倍体和双倍体之间的转换通过交配（单倍体到双倍体）和孢子形成（双倍体到单倍体）来实现。有两种单倍体细胞类型，分别为 a 型和 α 型。在一起生长时，这些细胞因交配而形成 a/α 双倍体细胞。在营养物匮乏的情况下，a/α 双倍体发生减数分裂（第 7 章），产生一个称为子囊的结构，每个子囊含有 4 个单倍体孢子（两个 a-孢子和两个 α-孢子）。但当生长条件改善时，这些孢子可以出芽并以单倍体细胞的形式生长或交配而重新形成 a/α 双倍体。

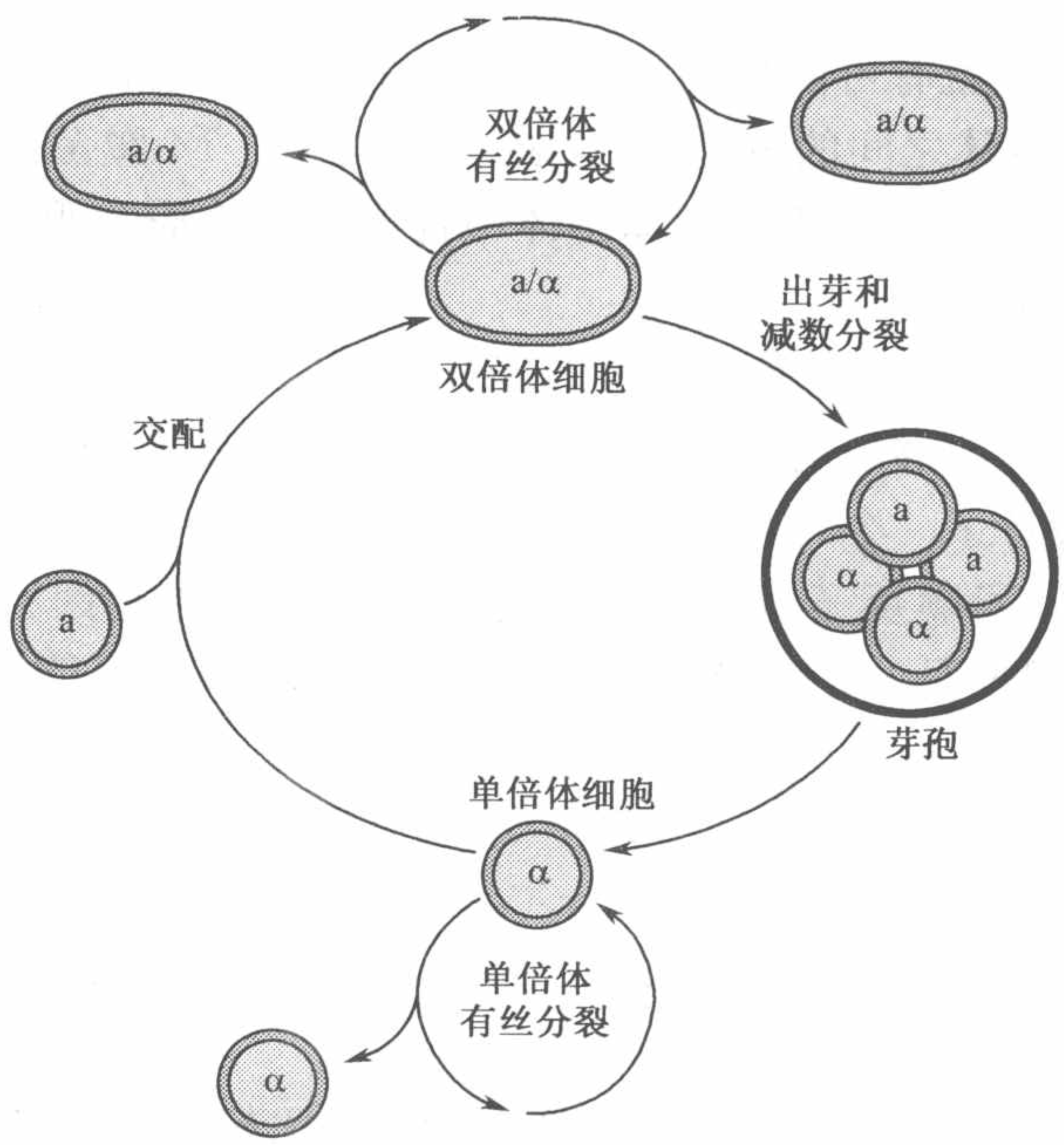


图 22-10 酿酒酵母 *S. cerevisiae* 的生活周期。如本章及其他章节所描述的，酿酒酵母以 3 种形式存在。两种类型的单倍体细胞，a 和 α，以及两者交配的双倍体产物。图示为这些不同种类细胞的复制、交配和芽孢形成。

在实验室中，可以用这些类型的细胞进行各种遗传分析。简单地将 2 个待测试的分别携带一个突变的单倍体菌株进行交配，就可以进行遗传互补实验。如果两个突变是互补的，相对于突变表型，双倍体就是一个野生型。如果要测试单个基因的功能，可以在只有一个拷贝的单倍体细胞里诱发突变。例如，想知道一个特定的基因是否是细胞生长所必需的，可以在单倍体里敲除这个基因。单倍体细胞只能承受非必需基因的敲除。

在酵母中容易产生精确的突变

精确和快速地改变单个基因的技术使酿酒酵母的遗传分析研究水平得到了进一步提高。末端与基因组的任何一个特定区域同源的线性 DNA 引入酿酒酵母细胞中，就会观

察到频率非常高的同源重组，导致相应染色体序列被转化所用的 DNA 片段取代（图 22-11）。利用这一特性可以对基因组做精确的改变。这种方法可以用来精确地删除整个基因的编码区、改变一个可读框中的特定密码子，甚至改变启动子中的一个特定碱基对。这种在基因组中进行精确改变的技术使研究特定基因或其调节序列的功能等具体问题变得比较容易。

酿酒酵母有一个研究透彻的小基因组

由于酿酒酵母遗传研究的历史很长，基因组又相对较小，所以它成了进行全基因组测序的第一个真核（非病毒）生物。这一“里程碑”式的事件完成于 1996 年。序列分析（1.3×10^6bp）发现了大约 6000 个基因，首次揭示了指导真核生物形成所需的遗传复杂性。

酿酒酵母的完整基因组序列的获得使酵母的研究可以在“全基因组”水平上进行。例如，包含来自所有近 6000 个酿酒酵母基因序列的 DNA 微阵列已广泛用于研究不同生理条件下基因表达模式的研究。实际上，目前已经对 200 多种不同条件下酿酒酵母细胞的基因表达水平进行了检测，包括不同的碳源（如葡萄糖对半乳糖）、细胞类型和培养温度。这些发现不仅对检测单个基因的表达非常有用，而且使我们能将基因按照对不同培养条件的应答分为各种协同作用的基因群。

其他全基因组资源还包括一个 6000 个菌株的文库，每一菌株中只删除了一个基因。其中的 5000 多株在单倍体状态时能够存活，表明大多数酵母基因是非必需的。这个酵母菌株库使检测酿酒酵母基因组的每个基因在特定过程中作用的遗产筛选得以进行。微阵列的使用还使利用染色质免疫沉淀技术进行转录因子的结合位点的全基因组作图得以开展（第 21 章）。

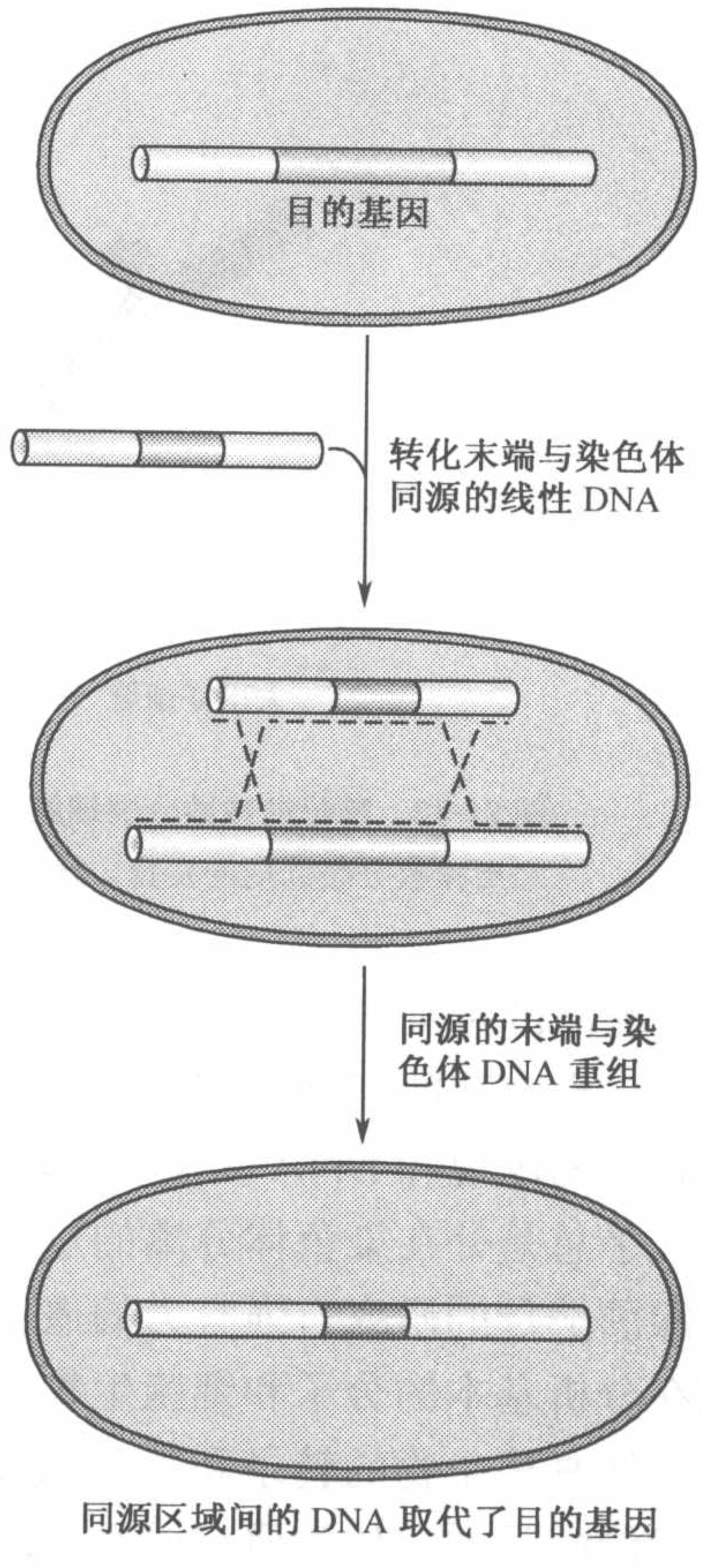


图 22-11 酵母中的重组转化。如文中所述，酵母基因组的所有区域都可以容易地被目标序列所取代。要插入的 DNA 片段的两端序列要与被取代的染色体区域的侧翼序列同源。当供体片段被导入细胞后，在这个生物中的高水平同源重组保证了与染色体的高频率重组，导致了所示的遗传交换。插入的 DNA 可能与原来的 DNA 只有一个碱基对的不同，也许是另一个极端，即长度和序列上非常不同。由此可以完成非常精细的遗传修饰。

酿酒酵母细胞在生长时改变形态

当酿酒酵母细胞按照细胞周期生长时，会发生形态上的特征变化（图 22-12）。当一个新细胞从它的母细胞中释放出来后，子细

胞稍呈椭圆形；随着细胞周期的延续，酿酒酵母细胞上形成了一个小“芽”，这个“芽”最终会成为一个独立的细胞。一直长到与其母细胞大致相当时，芽就会从母细胞中释放出来，两个细胞开始新的一轮繁殖过程。

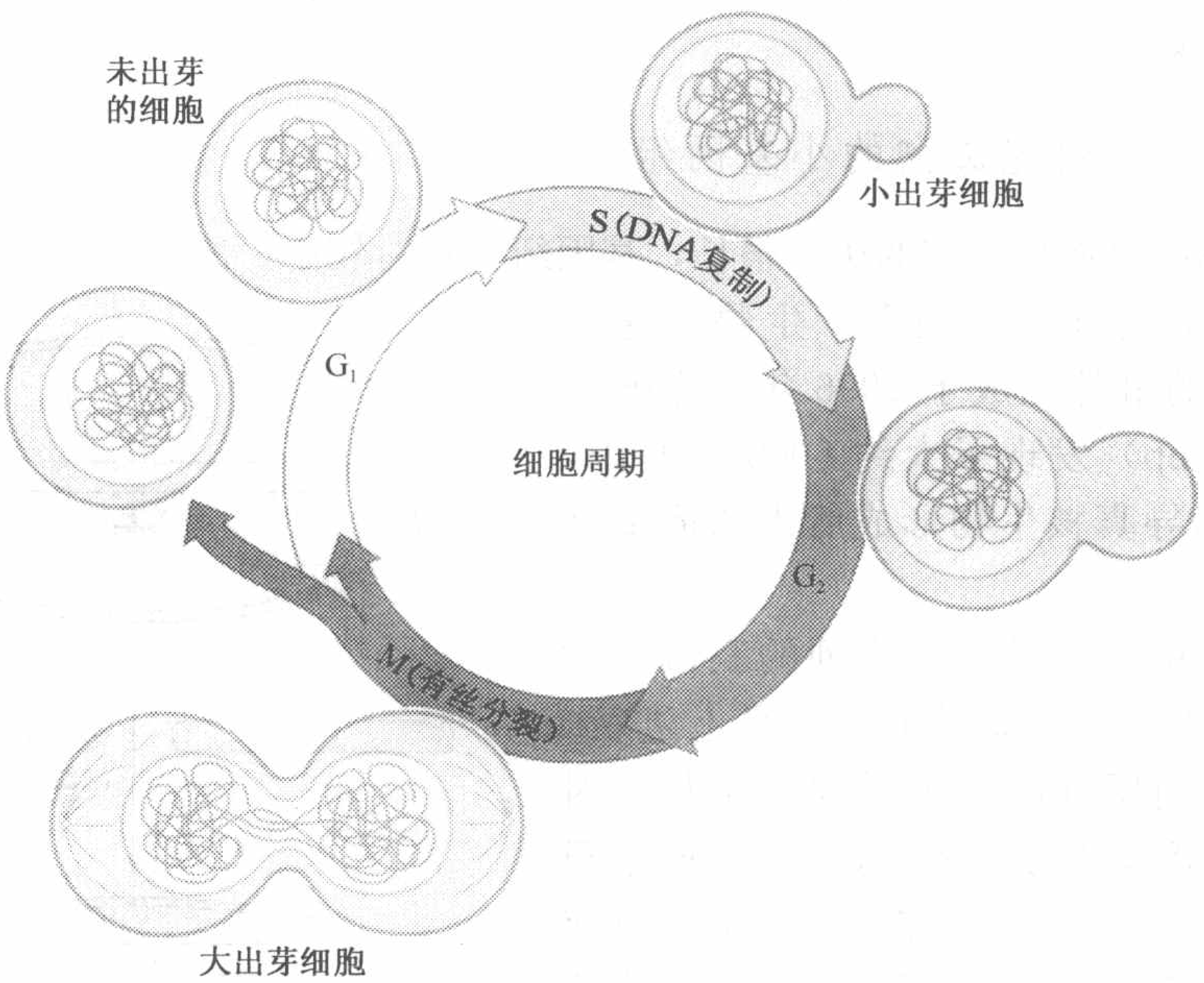


图 22-12　酵母的减数分裂细胞周期。酿酒酵母通过出芽进行分裂。图示一个子代芽经历减数分裂周期的发育，该过程在正文中也有描述。

对酿酒酵母细胞形状进行简单的显微观察，就可以获得许多有关细胞中发生的事件的信息。没有芽的细胞还没有开始复制它的基因组，这是因为在野生型酿酒酵母细胞中，新芽的出现和 DNA 复制的启动是紧密关联的。同样地，正在生长的、有着大芽的细胞几乎总是处在染色体分离的过程中。

酿酒酵母的研究有强有力的遗传学、生物化学和基因组学工具，这使酵母成为科学家们在分析基本的分子和细胞生物学问题时所喜爱采用的真核生物。酿酒酵母的研究为我们了解真核生物的转录和基因调控、DNA 复制、重组、翻译、剪接等做出了重要贡献。酿酒酵母的遗传学研究已经发现了所有这些事件中涉及的蛋白质。也许最为重要的是，在酿酒酵母中发现的那些对基本生命活动重要的蛋白质和基因在包括人类在内的其他真核生物中几乎总是保守的。因此，从这个简单真核模式生物中归纳的东西几乎总是与复杂生物的相同事件相关。

拟南芥

起源于农业和草药的植物科学在所有生命科学中历史最长。通过孟德尔的工作，植

物科学奠定了遗传学的基础。通过达尔文、Barbara McClintock、William Bateson 和其他人的工作，为细胞遗传学、发育学、生理学和进化科学奠定了基础。植物科学继续在基础领域如 RNA 干扰方面有着重要的进展，同时和过去一样，继续对经济和环境发挥着重要影响。在过去的几十年中，一种不起眼的类似芥末的杂草，拟南芥，渐渐成为与果蝇、秀丽新小杆线虫和小鼠媲美的模式生物。比如的动物同行更重要的是，拟南芥揭示了植物生物学，尤其是被子植物（开花的种子植物）的大多数关键问题。就像在 20 世纪玉米带来植物遗传学的革命一样，拟南芥注定要为植物基因组学和植物生物学的大多数问题做出贡献。

拟南芥具有单倍体和双倍体交替的快速生命周期

像酵母一样，所有植物的生命周期具有单倍体和双倍体期，二个时期分别根据它们的产物来命名——双倍体时期（与酵母类似）发生减数分裂并产生孢子，因此被称为“孢子形成体”（带有孢子的植物）。单倍体的孢子发芽产生单倍体期，在这个时期分化出两种性别的配子，所以单倍体期也称为雄配子体期或雌配子体期。在受精过程中，配子融合产生双倍体的合子。单倍体和双倍体期的相对长度随植物而异——苔藓植物大部分时期属配子体期，而拟南芥和其他高等植物大部分时间是孢子形成体，蕨类植物则介于两者之间。在开花植物中，配子体期非常短，只有 2 或 3 次有丝分裂，生殖细胞存在于成体植物长出的花中，而不像动物中那样隐藏的胚胎中。大多数植物（如拟南芥）是雌雄同体的，从分化的花中或发生雌雄减数分裂的花器官中产生两种性别的单倍体配子体（图 22-13）。

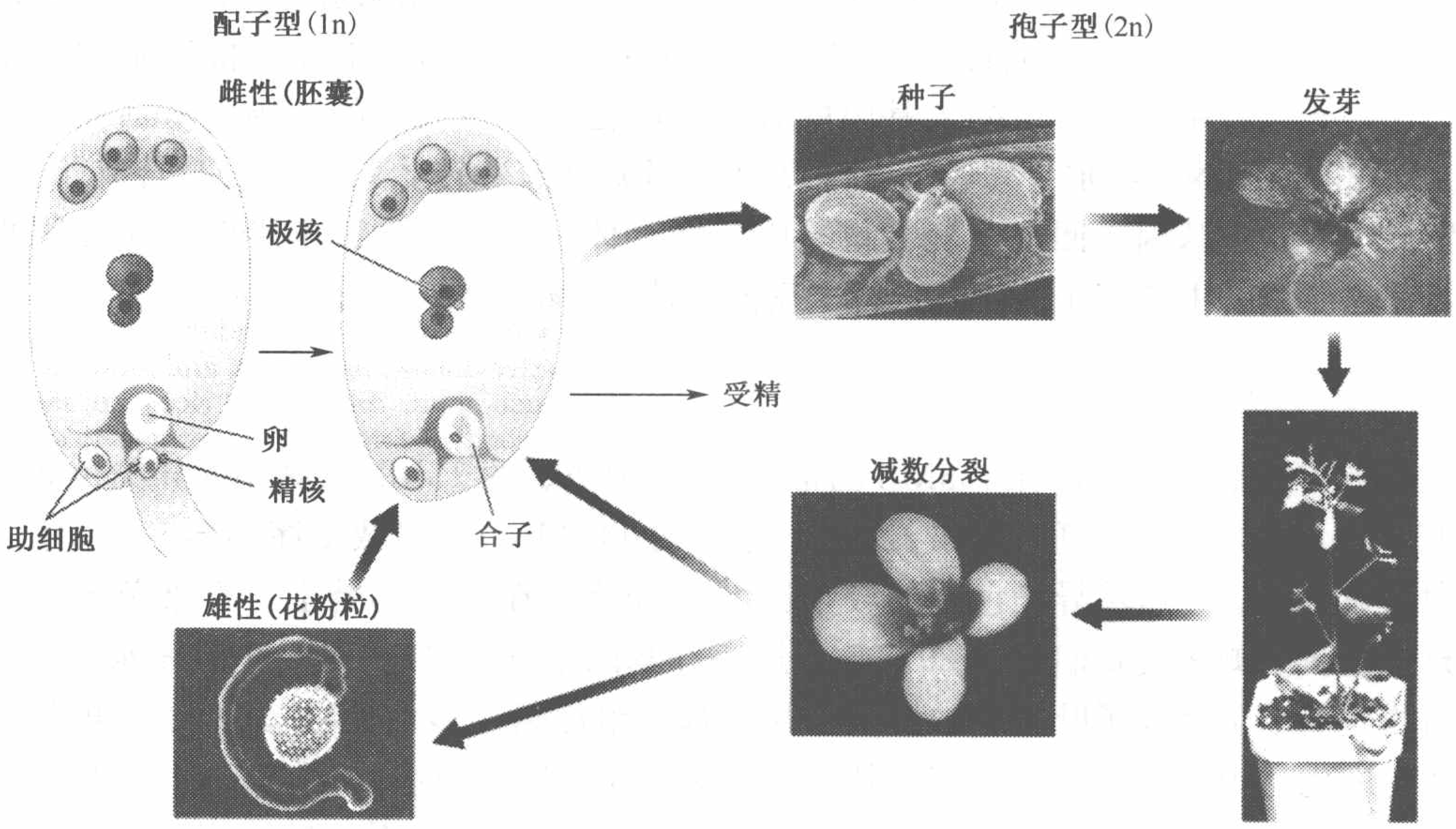


图 22-13 拟南芥的生命周期（Rob Martienssen 惠赠）。

种子植物，如拟南芥，在受精后经历另一个时期——胚胎发生和休眠，从而产生以

此命名的种子。同哺乳动物一样，胚胎由完成分化的胚胎外组织提供营养。但是与动物不同的是，这些胚胎外组织（称为胚乳“含在种子中”）是独立受精产物，它们的受精发生在第二个单倍体精子和雌性双倍体“中央”细胞，双倍体“中央”细胞则是由两个单倍体姐妹卵细胞融合而成。三倍体核快速分裂，随后又分裂形成很多细胞，并且快速积累淀粉和蛋白质，为胚胎提供重要的营养。由于胚胎发育时胚乳的营养被胚胎攫取，在许多植物中（如拟南芥）其生命是短暂的，但在其他植物中胚乳可以提供种子发芽所需的营养，而在主要谷物，如小麦和玉米，则可以提供淀粉。

拟南芥容易转化适于反向遗传学研究

土壤细菌根瘤农杆菌及其亲缘较近的细菌可感染植物，引发肿瘤性生长（冠瘿，galls），其原因在于细菌的 Ti（tumor-inducing）质粒携带着激素生物合成基因，该基因转移到了宿主植物的染色体 DNA 上。肿瘤引发（tumor-inducing）基因位于质粒的转移 DNA（T-DNA）部分，两个边界序列是转移所需的正向重复序列。将感兴趣的基因取代肿瘤引发基因，就可能转化植物。拟南芥的转化只需简单地在植物上喷上或将植物浸泡在含有促进感染的表面活性剂的高浓度农杆菌培养液中就行了。瞬时感染几乎在侵染的同时就可以发生，可用于瞬时表达研究。但稳定转化一般发生在数天或数周后，可以通过在受精前感染雌性配子。在 T-DNA 中加上选择性标记基因（抗不同的除草剂），并在含有除草剂的培养基上或土壤中发芽种子，就可以选择转化的植物。

拟南芥的转化效率很高，可以用来大量制备突变体。随机把几十万个 T-DNA 插入到植物中，然后对插入位点进行扩增和测序，已经收集到了很多带有插入基因的植物，插入事件发生在基因组的大多数基因中。这些插入突变可以用来进行与酵母中利用缺失所进行的相同的“反向”遗传学研究。进一步讲，在 T-DNA 中加上报道基因或增强子，这些插入可用来报告所插入基因的表达，或在通常不表达的细胞中激活基因的表达。通过在 T-DNA 中加上转座因子，就可能不需要额外转化就产生大量转座子跳跃，产生衍生的同位基因、回复突变等位基因及嵌合体。由于水稻和玉米的转化要困难得多，在这些植物中转座子一直是进行“反向”遗传学的主要工具。

拟南芥基因组小，容易操作

拟南芥的基因组只含有 105Mb 的常染色体 DNA、约 15Mb 已测序的异染色质和另外 15～25Mb 的卫星重复序列和 rDNA，总共约 140Mb。大多数测序的异染色质位于 5 个中心粒的两边，虽然异染色质的较小区域（knob）在染色体臂上。常染色体和大部分异染色质的测序使我们知道 29 000 个拟南芥基因中 99%的序列。许多其他植物基因组的测序揭示了真双子叶植物中发生了几轮基因组加倍（多倍性），真双子叶植物是被子植物进化树中的一个主要分支，其中包含拟南芥。最近的基因组加倍仅仅发生在几百万年之前，所以约 25%的拟南芥基因保留了一个功能同源基因，导致相当多的基因重复以致影响了反向遗传策略的应用。另一方面，也许部分是因为基因重复，在拟南芥中正向遗传 学一直非常有效，可以使用高剂量的诱变剂［ethylmethane sulfate（EMS），乙酰基亚硝基脲（ENS），或放射线］而不会杀死植物，相对数目较少的经诱变剂处理

过的种子就可以使突变达到饱和。种子可以直接进行诱变处理，隐性突变可以简单地通过让种子发芽后自变获得。

有了基因组序列和具有多态性的拟南芥品系后，对正向遗传学发现的突变进行定位克隆变得异常方便。EMS 突变甚至可用于反向遗传学策略，称为覆瓦（tilling），即筛选诱变剂处理过的植物的 DNA 来获得感兴趣的基因的点突变。随着高通量的测序方法的出现，这种策略甚至更具操作性，并且可以在每个基因中获得不同等位变异。基因组序列的测定使许多其他基因组学技术成为可能，如覆瓦微阵列、高通量蛋白质定位及蛋白质组学技术等。

通过小 RNA（19～30 个核苷酸）进行 RNA 干扰是一种重要的基因调控机制，无论是内源还是外源，该机制首先在植物中发现（第 18 章）。在拟南芥中，至少三类小 RNA——microRNA（miRNA）、反式短的干扰 RNA（tasiRNA）、与重复序列相关的 siRNA，它们在大小和生物起源方面不同，但都可以通过与目标序列配对并启动由核酸内切酶活性行使的“切割”、转录停止、染色质及 DNA 修饰等来调控基因。这些小 RNA 来源于单链“发夹”结构前体或双链 RNA，后者是依赖 RNA 的 RNA 聚合酶的产物。在拟南芥中采用 RNA 干扰的基因组方法包括 VIGS（virally induced gene silencing，病毒诱导的基因沉默）、共抑制、发夹沉默和人工 miRNA 等。

表观遗传学

表观遗传差异通常定义为可以通过染色体遗传，但不涉及核苷酸序列改变的“突变”。相对于通常的突变，这些“表观突变”的可逆频率要高得多，并且与 DNA 及相关蛋白质（尤其是组蛋白）的化学修饰有关。几十年来，植物一直处于表观遗传学研究的前沿，拟南芥也不例外。类似于哺乳动物基因组，但不同于酵母、线虫和果蝇的基因组，植物基因组内有胞嘧啶甲基化的高度修饰，这些修饰与组蛋白修饰一起对基因表达和基因组中发现的重复序列的表达具有表观遗传后果。这些修饰由各种各样的因子介导，包括首先在植物中发现的导致依赖 RNA 的 DNA 甲基化现象的 RNA 干扰，以及 RNA 依赖的组蛋白修饰，后者也在裂殖酵母和其他生物中发生。

当一个基因的沉默在雄性和雌性生殖细胞系中不同时，印记导致（通常）母本遗传的等位基因进行表达。被印记的基因表达在胚胎外的胚乳组织中普遍存在，使人想起在哺乳动物胎盘中的印记。在这些研究得很透彻的例子中，被印记的基因的去甲基化发生在中央细胞，导致胚乳中母本等位基因的表达。

表观遗传作用经常受环境的影响，在一个极端的例子中，植物在春天开花记忆的是冬天的寒冷。这种记忆是由寒冷来诱导、靠细胞的无性繁殖来保持、减数分裂又抹去这个记忆，导致大家习惯的类似冬麦的开花习性。在拟南芥中，这个过程（春化）是由 RNA 加工及组蛋白修饰来调控的，并且涉及 Polycomb 复合体，该复合体也参与动物的细胞记忆。

植物对环境的反应

与动物不同，植物固定在一点上，不能逃避环境的攻击，导致一些通常在动物中没

有的特性——如对啃食的耐受性。植物的天然免疫系统，最早在分子水平上被描述为“基因对基因”反应，涉及许多在动物中保守的基因，但它们是高度多样化的，可以识别病毒、微生物、线虫、昆虫，甚至其他植物。除了这些“生物的”逆境外，植物必须经受住“非生物的”逆境并对其做出反应，这些逆境包括光强度、生物节律、营养、盐和水的改变等。许多环境因素对发育会有深刻的影响——如通过诱导或延迟开花使之更适合种子产生。

光在植物生物学中扮演一个中心的角色，进行光合作用的叶绿体来源于古代的共生原核生物，负责固定生物圈中大多数有机碳。由于在遗传和基因组操作方面的方便性，甚至在光合作用研究领域，拟南芥也正在取代传统的生理学模式生物——如烟草和菠菜。

发育和形态建成

同植物生物学的其他领域相比，植物发育对作物驯化和育种的影响最大，所以对人类历史的影响也最大，其中由古代农民及富有经验的育种家等进行的选择是巨大的创新，他们改进了花序结构、种子落粒性、叶片形状等。与现代农民视为草的祖先物种相比，花椰菜、爆花玉米和羽衣甘蓝与他们只有几个基因的差异。由于开花植物是最近进化的群体，许多负责开花的基因是用拟南芥作为模式生物被发现的。

一般而言，植物和动物是从一个共同的单细胞祖先分支进化而来，所以多细胞发育是在各自的界中独立进化的。因此我们看到通用的基本原则，如转录因子和信号传导层次结构（多肽、激素、和受体）的核心地位，在每个界中都遵守，而特异的分子就很少保守。有些机制非常相似，如细胞周期和 MAP（mitogen-activated protein）激酶级联途径，但大多数是明显不同的。例如，在两类谱系中同源异型和异时特征都是由转录因子和 miRNA 保证的，但涉及的分子不同，在植物中大多是 MADS（MCM1，agamous，deficiens，and serum response）转录因子，在动物中是 *Hox* 基因。在两个界中细胞间交流都涉及激素，但这只能说是类似（植物和动物都使用类固醇也许是个例外）。事实上，植物中的高度关联的超细胞维管系统允许大分子在细胞间通过，如 mRNA、小 RNA 和转录因子本身，而这个现象在动物中很少见到。

那么，共同的发育机制可能在单细胞或寡细胞祖先中发挥功能，并且可能被指派去独立地执行相似的功能。也许在祖先的单细胞真核生物保守的表观遗传学机制，如 Polycomb 系统，在基因组的组织、基因组防御、染色体生物学和细胞分化方面执行功能，而不是多细胞转录记忆。拟南芥对发现植物界以及与动物界间保守功能方面扮演着主角。

秀丽新小杆线虫

在分子遗传学领域做出杰出贡献后，Sydney Brenner 又确定了一种适合于研究发育重要问题和行为分子基础的很小的多细胞动物。他吸取了噬菌体和细菌的分子遗传学研

究的成功经验，希望尽可能采用最简单的、既具有分化的细胞类型又符合类似微生物遗传学的生物。1965 年，他决定用秀丽新小杆线虫（*C. elegans*），因为它具有许多合适的特性。例如，快速的世代时间使遗传筛查易于进行；雌雄同体的自身繁殖可以产生几百个“自身后代”，因而适于大量产生子代生物体；有性繁殖使遗传品系可以通过交配来构建，以及少量透明细胞的存在使发育过程能被直接追踪。

Brenner 设定了两个最终实现将秀丽新小杆线虫用作模式生物至关重要的雄心勃勃的初始目标。目标之一是通过重建系列切片电镜图绘制所有细胞的完整图谱（1986 年由 John White 完成）；目标之二则是绘制细胞谱系（1983 年由 John Sulston 完成）。7 年后 Brenner 分离到了 300 多个含有形态和行为突变的突变体，建立了这个新的模式生物的遗传学。这些突变体分为 100 多个互补组，定位于 6 个连锁群。将近 30 年以后，全世界有 400 家实验室在研究秀丽新小杆线虫。由于秀丽新小杆线虫简单而容易用于实验，它是目前研究得最透彻的多细胞动物之一。

秀丽新小杆线虫的生活周期很短

秀丽新小杆线虫在培养皿里培养，饮食很简单，只吃细菌。它们在一定的温度范围内生长良好，在 25℃比在 15℃生长速度快一倍。在 25℃，受精的胚胎在 12h 内完成发育，孵化成自由生活的动物，能够进行复杂的行为。第一个幼虫期（L1）分成 4 个时期（L1～L4），在 40h 内成为一个性成熟的成虫（图 22-14）。

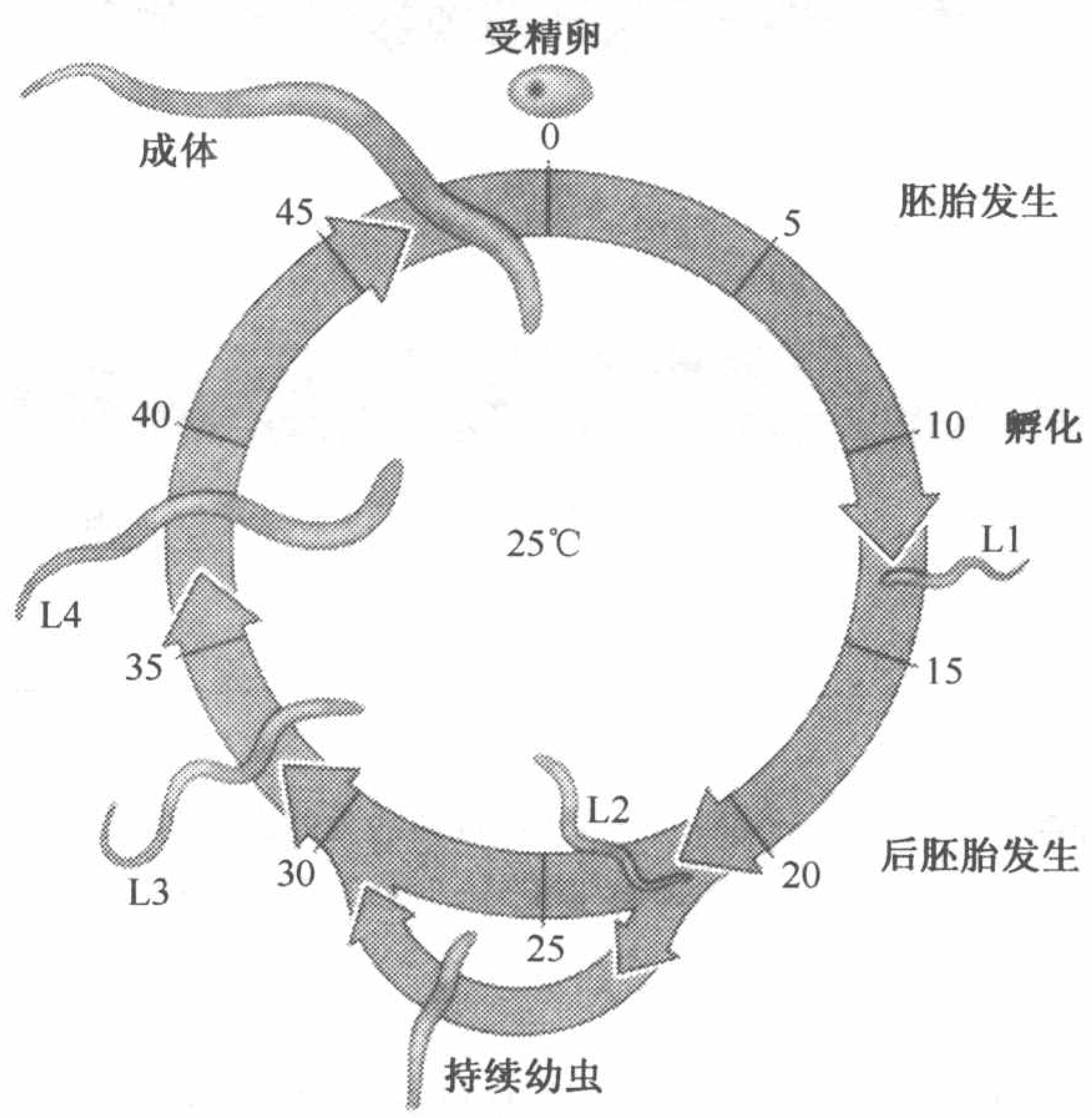


图 22-14　秀丽新小杆线虫的生活周期。所示为生活周期，如文中所述，从第一期幼虫到成虫的发育以小时计。L1 幼虫进入的另一个发育期（成为持续幼虫）也在此显示。

雌雄同体的成虫大约 4d 内就可以产生多达 300 个自身后代，或与罕见的雄性线虫

交配，产生多达 1000 条杂交后代。成虫存活大约 15d。在有生长压力的条件下（少食、高温、高密度），L1 期动物可以进入另一个发育期，形成一个所谓的持续幼虫（dauer）。持续幼虫对环境压力有抵抗能力，能存活几个月的时间等待环境条件的改善。通过研究那些不能进入或不适当地进入持续幼虫期的突变体，发现了特异性神经元中表达的感知环境条件的基因、在动物全身表达的调控身体生长的基因，以及控制寿命的基因。激活控制寿命的基因可以戏剧性地延长动物的寿命，这些基因的同源基因被认为与哺乳动物的寿命延长紧密相关。

秀丽新小杆线虫由相对较少、研究得很深入的细胞谱系组成

秀丽新小杆线虫身体构造简单（图 22-15）。成虫两性体的重要器官是性腺，其中含有正在增殖和分化的生殖细胞（精子和卵子）、受精室（受精囊），以及暂时储存幼小胚胎的子宫。胚胎经过阴门从子宫排到体外，阴门是一个由 22 个表皮细胞组成的结构。妨碍阴门形成的突变并不影响胚胎的产生，但却使排卵无法进行。结果，胚胎在子宫内发育并孵化，孵化了的线虫随后吞食它们的母亲，并被限制在她的皮肤里（表皮层），形成一个“线虫袋”。这个很容易确认的表型已经使几百个没有阴门的突变体得到分离，从中发现了许多控制阴门细胞的产生、特化和分化的基因。其中包括高度保守的受体酪氨酸激酶信号传递途径的成员，该途径控制细胞增殖。

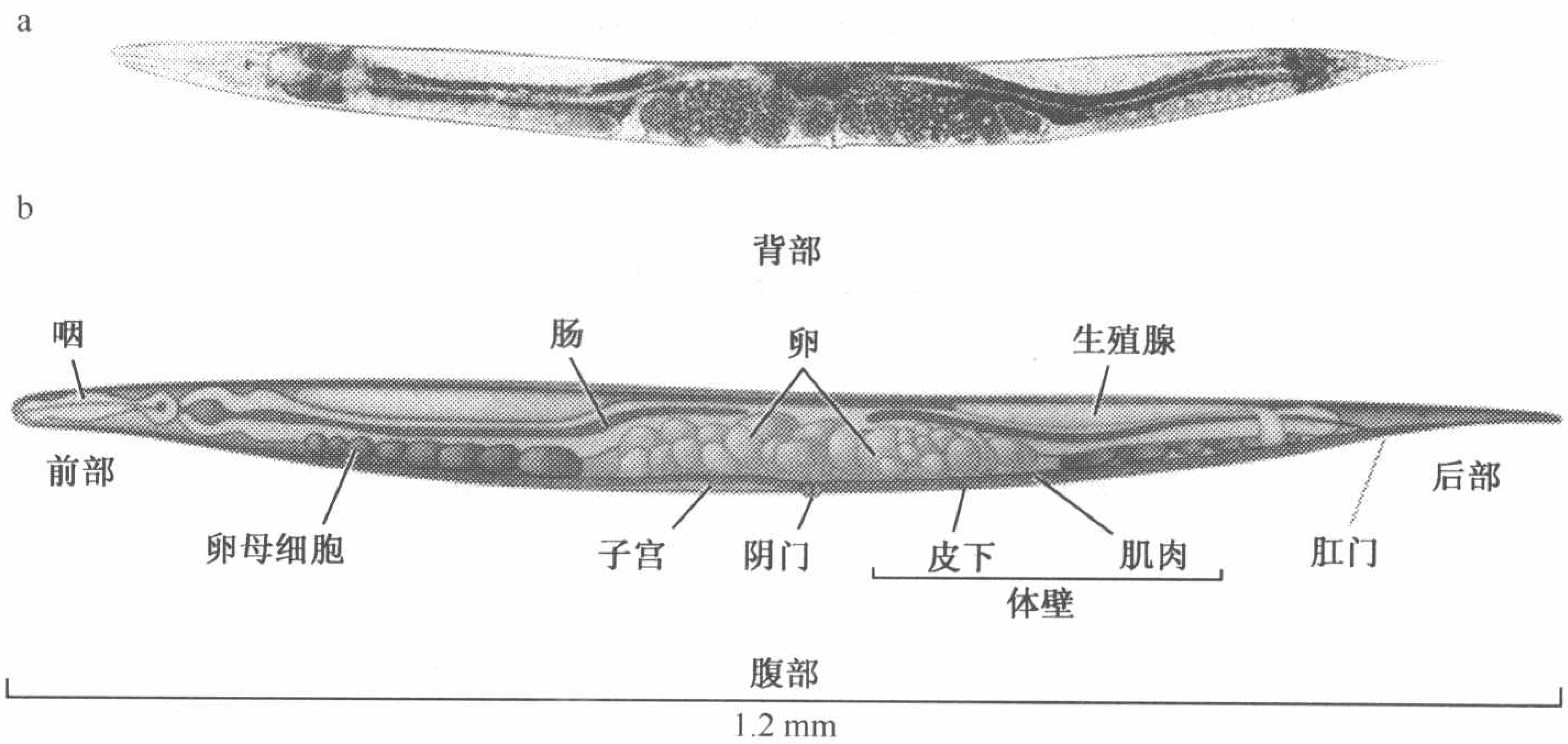


图 22-15 线虫的身体平面图。上面 a 显示的是一个成虫两性体线虫的纵切剖面图。各种器官在下面的略图 b 中标示，并在文中有所描述。［a，经许可，影印自 Sulston J. E. 和 Horvitz H. R. 1977. *Dev. Biol.* 56：110-156. © Elsevier.］

在哺乳动物中，这些基因的许多同源基因是致癌基因和肿瘤抑制基因，当它们发生改变时就会导致癌症。在秀丽新小杆线虫，使这个途径失活的突变体不会产生阴门细胞，所以阴门不能发育，而激活这个途径的突变体则导致阴门前体细胞的过度增殖，产生具有多个阴门的表型。因为线虫是透明的，阴门仅仅来源于 22 个细胞，就可能以细胞水平的分辨率描述突变体的缺陷，从而将突变的类型与某一特定细胞的改变联系起来。

细胞死亡途径是在秀丽新小杆线虫中发现的

到目前为止，秀丽新小杆线虫研究中最引人注目的成就是阐明了调节细胞凋亡或细胞死亡的分子途径。细胞谱系的早期分析注意到在每一个动物中成组细胞的死亡都是相同的，暗示细胞死亡是受遗传控制的。最初分离到的细胞死亡缺陷型（*ced*）突变体是那些在相邻细胞吞噬死亡细胞残体方面有缺陷的突变体。因而，在突变体内，死亡细胞残体会存留多个小时。利用这些 *ced* 突变体，H. R. Horvitz 和他的同事分离了更多的不能产生保留细胞残体的 *ced* 突变体。这些突变体被证明是在启动细胞死亡程序方面有缺陷。除了一个特例外，*ced* 突变体的分析表明，所有发育中的程序化细胞死亡是自主的，也就是说，是细胞自杀。在雄性中，有个叫做连接细胞的细胞被它的相邻细胞杀死。*ced* 基因的鉴定为确认哺乳动物的同类蛋白质提供了方法，这类蛋白质在所有的动物中基本上驱动同样的生化反应来控制细胞死亡，事实上，在秀丽新小杆线虫里表达人类的 *ced* 同源基因可以替代突变的 *ced* 基因。细胞死亡在发育和疾病过程中与细胞增殖同等重要，并且是许多控制癌症和神经退行性疾病治疗方法的焦点。

RNA 干扰（RNAi）是在秀丽新小杆线虫中发现的

1998 年，一项非同寻常的发现被公布：双链 RNA（dsRNA）引入秀丽新小杆线虫后，抑制了与该 dsRNA 同源的基因的表达。这个意外的发现以及接下来的 RNA 干扰（RNAi）分析具有两个方面的重要意义。第一，RNAi 看来是普遍存在的，因为几乎所有的动物、真菌或植物细胞中 dsRNA 的引入都会导致同源 mRNA 的定向降解。实际上，我们对 RNAi 的了解很多来自于植物研究（第 18 章）。第二点是对这个神秘过程的实验研究揭示了它的分子机制（图 18-5）。这些研究与另一个 RNA 介导的基因调节过程相交叉，那个过程涉及微小的内源小 RNA，已有研究显示它们在植物和动物中调节基因表达、在纤毛虫里协调基因组重排并在酵母里调节染色质结构。最初的两个小 RNA 是在秀丽新小杆线虫中进行遗传筛选时发现的。部分线虫小 RNA 在两翼昆虫和哺乳动物中是保守的，它们的功能正在被逐步揭示。最近的研究表明人类基因组可能含有大约 1000 个 miRNA 基因。

黑腹果蝇

果蝇作为模式生物用于遗传和发育生物学研究已近百年。1908 年托马斯·亨特·摩根（Thomas Hunt Morgan）和他在哥伦比亚大学（Columbia University）的研究助手把腐烂的水果放在舍莫宏大厅（Schermerhorn Hall）的实验室窗户架上，目的是搜寻一种小的、快速繁殖的动物，并能够在实验室里培养来研究数量性状的遗传，如眼睛的颜色。在捕获的各种动物中，果蝇最符合要求。成虫在两星期中就可以产生大量子代，果蝇培养用的是回收的牛奶瓶和廉价的酵母、琼脂混合物。

果蝇的生活周期很短

果蝇生活周期的突出特征是有一个快速的胚胎形成期，然后是 3 个幼虫生长期及变形期（图 22-16）。胚胎形成在受精后的 24h 内完成，最终孵化出第一龄幼虫。如同我们在第 18 章中所讨论的，果蝇胚胎发育早期的核分裂是所有动物中最快的。一个第一龄幼虫生长 24h，然后蜕皮成一个更大的第二龄幼虫。这个过程重复以产生一个第三龄幼虫，它进食并生长 2～3d。

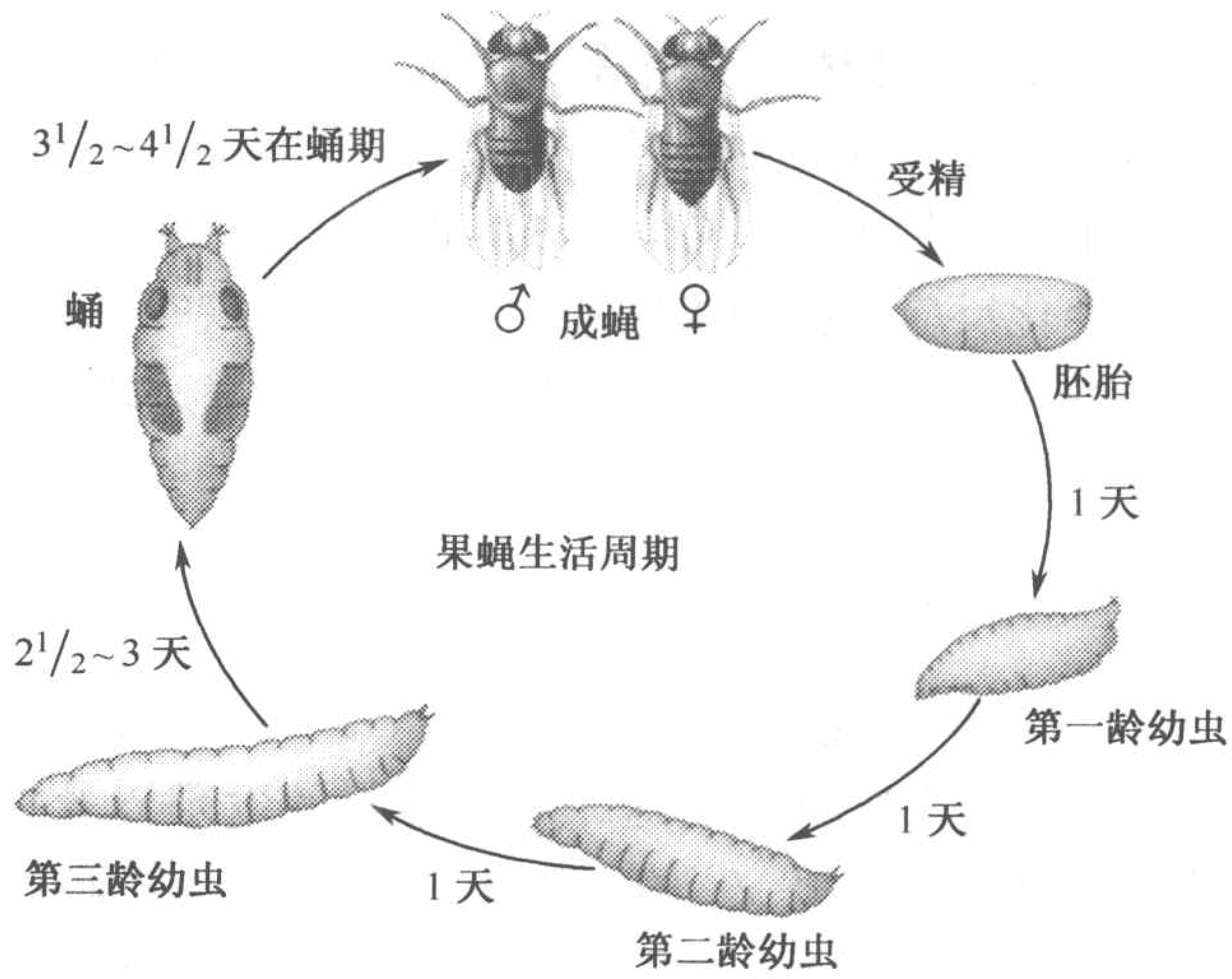


图 22-16 果蝇生活周期。图示为文中描述果蝇发育的各个时期。

幼虫发育中的一个关键过程是器官芽的生长，它是由胚胎中期表皮内陷形成的（图 22-17）。每套肢体都有一对器官芽（例如，一套前腿器官芽和一套翅膀器官芽）。还有眼睛、触角、口器和生殖器的器官芽等。起初器官芽很小，在胚胎中仅由不到 100 个细胞组成，但在成熟的幼虫中却含有上万个细胞。翅膀器官芽的发育已经成为理解诸如 Hedgehog 和 Dpp（TGF-β）等分泌信号分子的梯度是如何控制复杂的身体结构形成过程的重要模式系统。器官芽在变形（或化蛹）过程中分化成各自相应的成虫结构。

第一个基因图谱在果蝇中产生

1910 年，摩尔根实验室发现了一个自发突变的雄蝇，它的眼睛是白色的而不是正常品系中的亮红色。由这个果蝇启动了一系列深入的遗传学研究，产生了两项重要发现：一是基因位于染色体上，二是每个基因是由两个等位基因组成的，这两个基因在减数分裂过程中独立分配（孟德尔第一定律；第 1 章）。其他突变的发现证实了位于不同染色体上的基因独立地分离（孟德尔第二定律），而那些连锁在同一条染色体上的基因则不同。

哥伦比亚大学的学生 A. H. Sturtevant（摩尔根实验室的成员）设计了一个简单的数学运算方法，可以根据重组频率绘制连锁基因之间的距离。到了 20 世纪 30 年代，

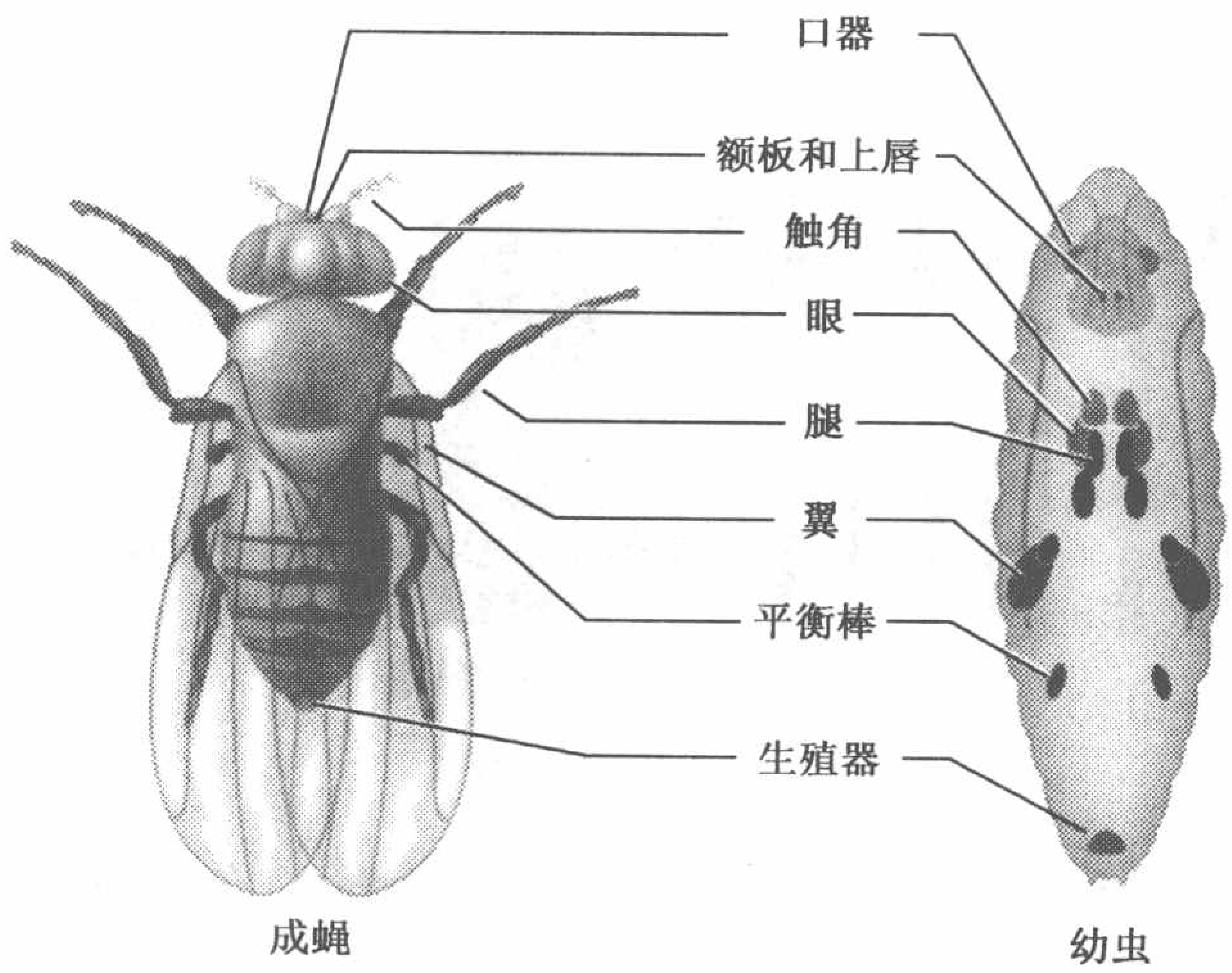


图 22-17 果蝇的器官芽。右图所示为幼虫中各个器官芽的位置。左图为它们在成虫中形成的肢翼和器官。这些器官芽起初在胚胎里只是很小的细胞群，但在成熟的幼虫里有上万个细胞。这些器官芽在化蛹的过程中发育成它们相应的成虫中的结构。

遗传图谱已经非常密集，确定了控制成虫多种物理特性的多个基因的相对位置，如翅膀大小、形状以及眼睛的颜色和形状等。

另一位曾在摩尔根果蝇实验室接受培训的科学家 H. J. Muller，第一次提出了环境因素，如离子辐射，导致染色体重排和遗传突变的证据。日常进行的大规模“遗传筛选”是通过喂食成年雄蝇一种诱变剂如 EMS（ethylmethanesulfonate），然后将之与正常雌蝇交配。F_1 代是杂合的，含有一条正常的染色体和一条随机突变的染色体。研究这些突变的方法有许多种，详见以下描述。

除了强大的生育能力（一个雌性可以产生几千个卵）和快速的生活周期外，果蝇还有几个非常有用的特征，使之在实验研究中长期扮演重要角色。它只有 4 条染色体：2 条大的常染色体，即染色体 2 和 3，一条较小的 X 染色体（它决定性别），和一条非常小的第 4 条染色体。Muller 的同事 C. B. Bridges 发现在果蝇幼虫里某些组织中能够进行多次内部复制而并不发生有丝分裂。在唾液腺中，这个过程产生了极其巨大的由大约 1000 个拷贝的染色单体组成的染色体。Bridges 利用这些多线染色体（polytene chromosome）来绘制果蝇基因组的物理图谱（这是在所有生物中第一个绘出的）（图22-18）。

Bridges 在 4 条染色体上总共发现了大约 5000“条带”，并将许多条带和经典重组图谱遗传位点的位置建立了相关性。例如，隐性的 *white* 突变基因杂合的雌性果蝇表现出正常的红眼。然而，含有 *white* 突变的相似雌蝇若同时在另一条 X 染色体上移走了 3C2-3C3 多线条带则表现为白眼。这是因为已不存在该基因正常的显性拷贝。这类分析使我们得出“白眼基因位于 X 染色体的多线条带 3C2 和条带 3C3 之间”的结论。

其他各种遗传方法的建立，确立了果蝇作为动物遗传研究首要模式生物的地位。例如，制造出了平衡染色体（balancer chromosome），它含有一系列相对于原始染色体的

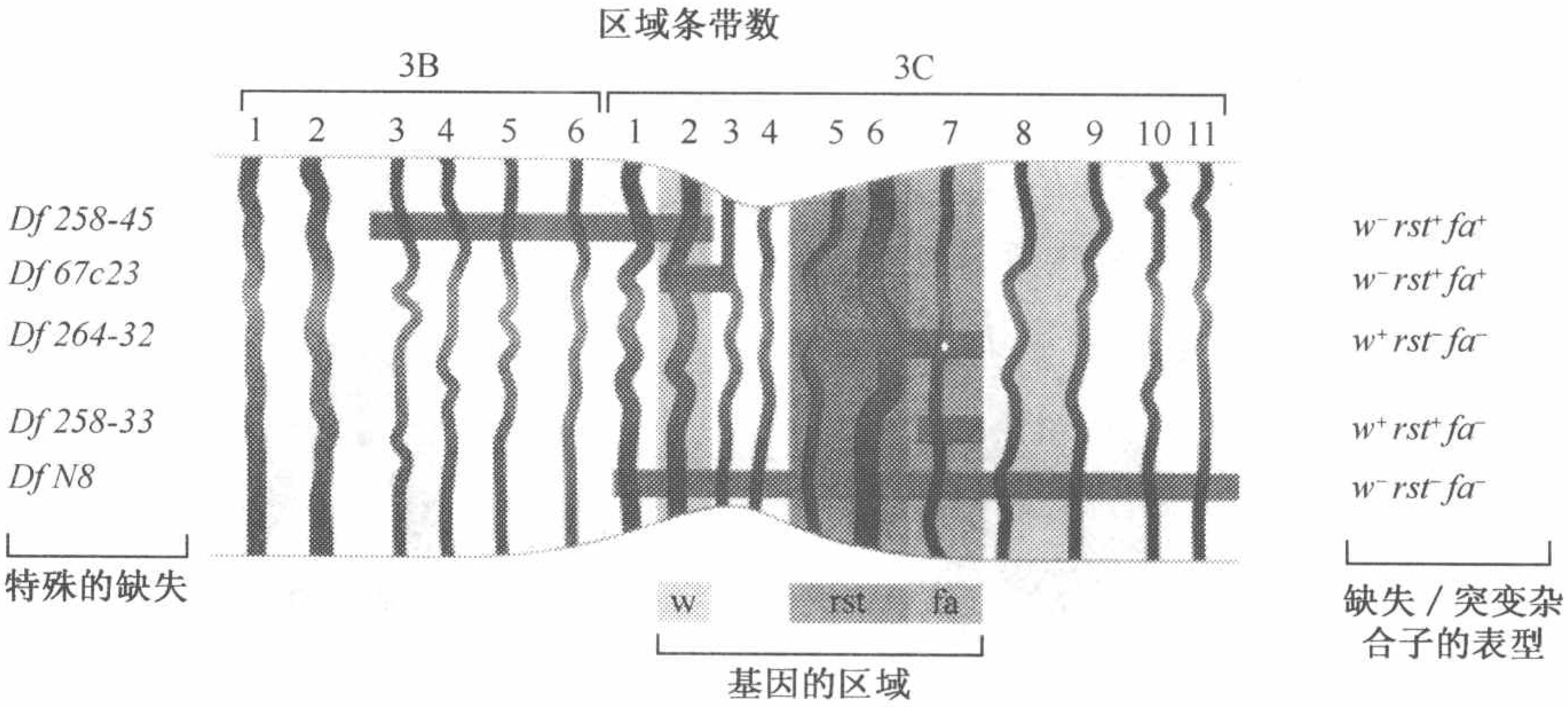


图 21-18 遗传图谱、多线染色体和缺失图谱（deficiency mapping）。在某些果蝇组织中，没有伴随有丝分裂的核内复制导致放大的染色体，最明显的是唾液腺体，在那里可以产生巨大的由 1000 个染色单体组成的染色体。这是第一次有可能把某些性状的基因出现与染色体的特定物理片段联系起来。具体地说，果蝇的表型（白眼）与染色体的缺失相关联（经许可，摘自 Hartwell L. et al. 2004. *Genetics*：*From genes to genomes*，2nd ed.，p. 816，Fig. D-4. © McGraw-Hill.）。

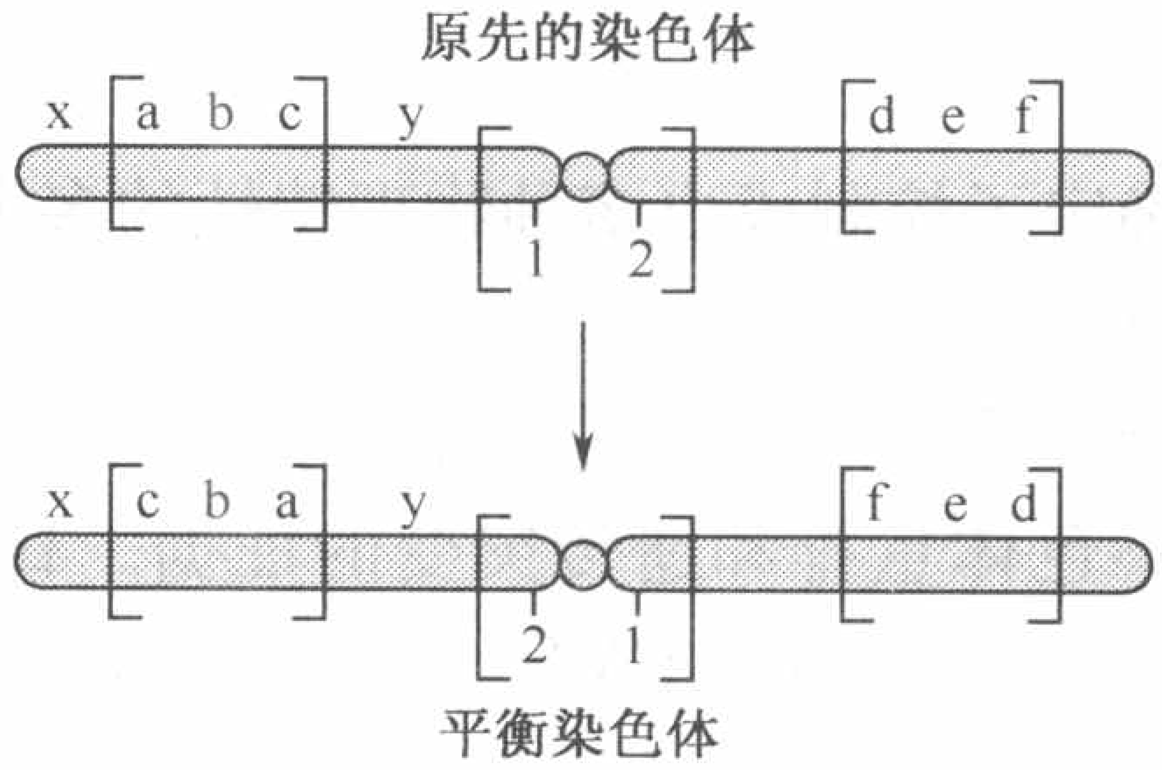


图 22-19 平衡染色体。当与原先的亲本染色体（上图）比较时，平衡染色体（下图）有一系列倒位。本图中，绘出的染色体有两个臂。平衡染色体的左臂有一个内倒位，基因 a、b、c 的顺序与原来相反。同样地，平衡染色体右臂也有一个倒位，颠倒了基因 d、e 和 f 的顺序。另外，以着丝粒为中心可能有一个倒位，此处颠倒了基因 1 和 2 的顺序。与原来的染色体相比，平衡染色体的基因顺序有明显不同。这样，含有一个正常拷贝和一个平衡染色体拷贝的杂合体中染色体之间的重组就被抑制了。

倒位（图 22-19）。严格地说，这种平衡染色体在减数分裂过程中不能与原始的染色体进行重组，就可能永久保留含有隐性致死突变的果蝇群落。我们在第 18 章讨论过的 *even-skipped*（*eve*）基因的一个无功能突变，这个突变纯合体的胚胎会死亡，不能产生幼虫和成蝇。*eve* 位点被定位在染色体 2（在多线条带 46C）上。这个无功能突变可以在杂合群体中保留，杂合子含有一条有 *eve* 无功能等位基因的“正常”染色体和一条含有该基因正常拷贝的平衡染色体。既然 *eve* 的无功能等位基因是完全隐性，这些果蝇是可以存活的，然而，在成体后代中只观察到杂合子的存在。含有两个平衡染色体拷贝的胚胎不能存活，因为某些倒位破坏了生存必需基因而产生隐性致死基因。另外，含有两个拷贝正常染色体的胚胎也不能存活，因为它们对 *eve* 的无功能突变是纯合的。

遗传嵌合体使分析成蝇中的致死基因得以进行

嵌合体（mosaics）是在一个总体上“正常”的遗传背景下含有小部分突变组织的动物。既然该生物中大多数的组织是正常的，一小部分的不正常不会致死。例如，通过X射线照射发育中的幼虫，诱导有丝分裂重组可以产生小块的 *engrailed/engrailed* 纯合突变组织。当突变部分出现在发育中的翅膀后面的区域，那么果蝇就表现出非正常的翅膀，在正常的后部结构区产生重复的前部结构。遗传嵌合体分析提供了Engrailed是将果蝇的附器和节片细分成前后两部分所必需的第一个证据。

最令人惊奇的遗传嵌合体是雌雄嵌合体（图22-20），这些果蝇是严格意义上的半雄半雌果蝇。果蝇的性别是由X染色体的数目决定的，有两条X染色体的个体是雌性，只有一条X染色体的是雄性（与小鼠和人类不同，Y染色体在果蝇中不决定性别，Y只在产生精子中需要）。在极为罕见的情况下，随着精子和卵子前核的融合，新受精的XX胚胎里两条X染色体中的一条会在第一次有丝分裂时丢失。

这种X染色体的不稳定性仅出现在第一次分裂时，在随后的分裂中，含有两条X染色体的细胞核产生两条X染色体的子核，而只有一条X染色体的核产生的子代含有一条X。如同我们在第19章所讨论的，这些细胞核经历快速分裂而细胞膜不分裂，然后迁移到卵的外围。这个迁移是连贯的，含有一条染色体的核和含有两条染色体的核很少或不发生混杂。因而，一半的胚胎是雄性的，一半是雌性的，尽管分开雄性和雌性组织的“分界线”是随机的，它的确切位置取决于在第一次分裂后两个子细胞核的方向。这条分界线有时把成蝇切成左半部是雌性的，右半部是雄性的。假设一条X染色体含有隐性的白眼基因，如果野生型X染色体在第一次分裂时丢失，那么果蝇的右半部，即雄性部分为白眼（雄性半部只有突变的X染色体），而左半部（雌性一边）为红眼（记住雌性半部有两条X染色体，携带显性的、

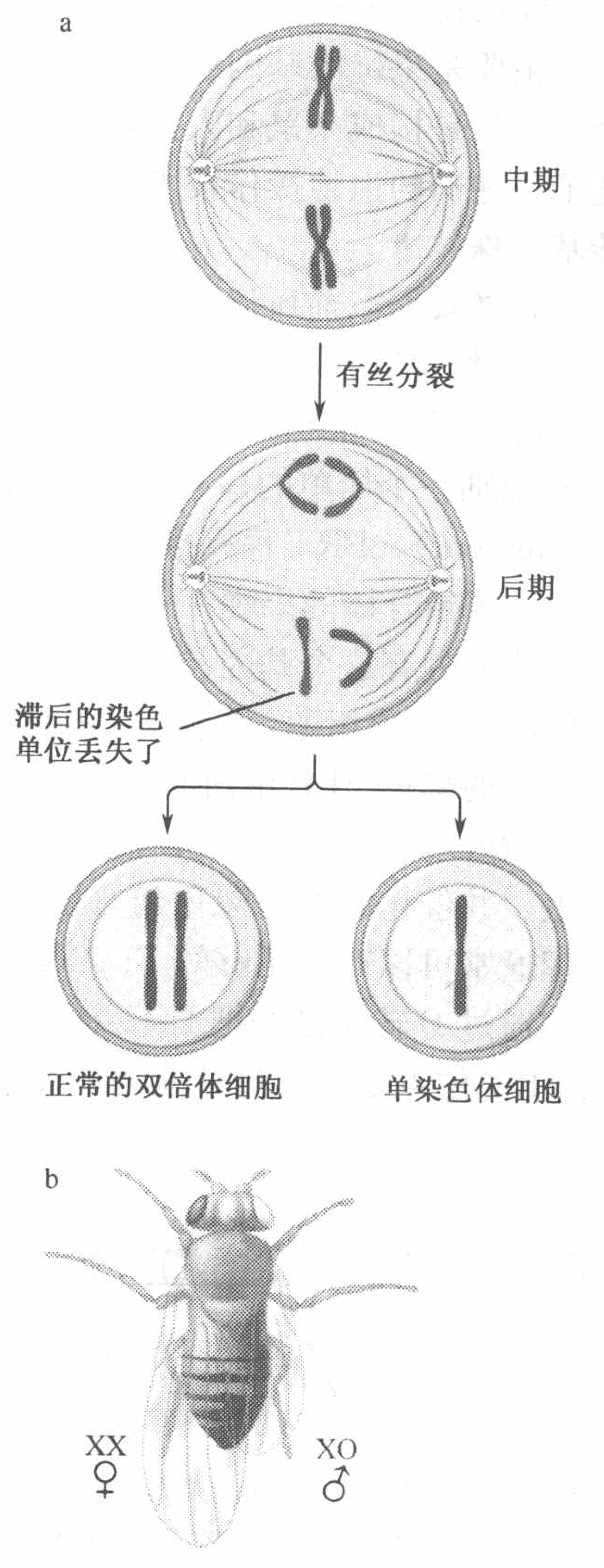


图22-20　雌雄嵌合体。雌雄嵌合体是一令人惊奇的遗传嵌合体的形式。a. 蓝色的X染色体携带隐性（白眼）突变，而红色的X染色体有该基因的正常的显性拷贝。如文中所述，突变是一个XX（雌性）果蝇的一条X染色体在第一次有丝分裂时丢失的结果。b. 在形成的突变体中，果蝇的一半是雌性，另一半是雄性。

野生型的等位基因)。

酵母 FLP 重组酶能够有效产生遗传嵌合体

经典遗传分析没有预料到的是，果蝇有若干个对分子研究和全基因组分析有利的特性，主要是果蝇的基因组相对较小，它只有大约 150Mb，蛋白质编码基因还不到14 000 个，这只是构成小鼠和人类基因组的 DNA 总量的 5%。果蝇进入当今时代后，人们改进了一些旧的遗传操作技术，建立若干全新的研究方法，如制备携带重组 DNA 的稳定转基因株技术。

就像我们以前所讨论的，遗传嵌合体是在体细胞里通过有丝分裂重组而产生的。起初，X 射线被用来诱导重组，虽然这个方法效率较低，只能产生小范围的突变组织。最近，酵母 FLP 重组酶的使用使有丝分裂重组的频率得到了很大的提高（图 22-21）。FLP 识别一个简单的序列基序，FRT，然后催化 DNA 重排（第 11 章）。通过 P 因子（P-element)，FRT 序列插入到 4 条染色体中每条染色体的着丝粒附近（见下）；随后就产生了杂合体的果蝇，一条染色体上含有 *Z* 基因的一个无功能等位基因，而其同源染色体上带有这个基因的野生型拷贝，两条染色体都含有 FRT 序列。因为在果蝇里没有内源 FLP 重组酶，所以这些果蝇是稳定存活的。然而，在含有热诱导 hsp70 启动子控制下的酵母 FLP 蛋白质编码序列的果蝇转基因株里，还有可能引入重组酶。一旦进行热激处理，FLP 就在所有细胞内合成。FLP 结合到两个含有 *Z* 基因的同源的 FRT 基序上，并催化有丝分裂重组（图 22-21)。这个方法非常有效。事实上，短时间的热激处理通常可以产生足够的 FLP 重组酶而在一个成蝇的不同区域产生大范围的 z^+/z^- 组织，通过 P 转座已经把 FRT 插入果蝇基因组各处。现在用 FLP 重组酶在目的基因两边的 FRT 位点之间重排就能制造几乎任何基因的缺失。

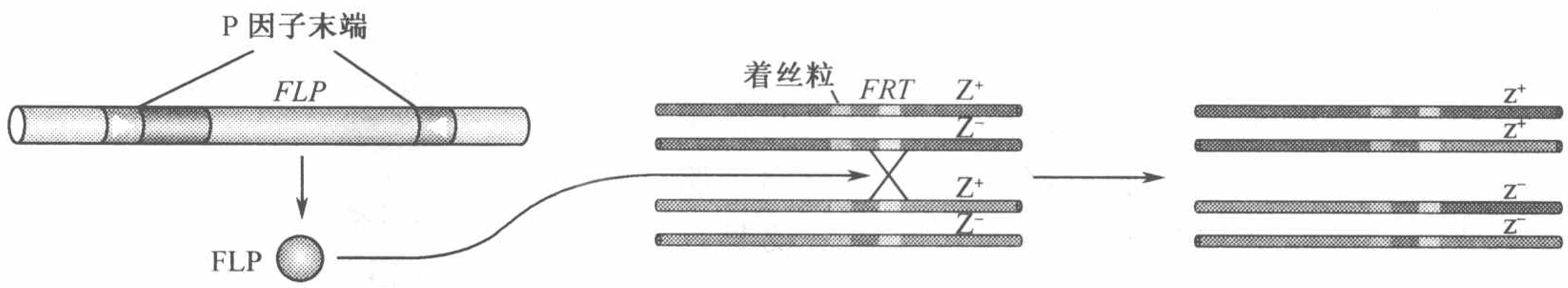


图 22-21 FLP-FRT。来源于酵母的定点重组系统（在第 11 章中有描述）的使用促进了果蝇中高水平的有丝分裂重组。重组的控制是通过在需要的时候在果蝇中表达重组酶而实现的。

制备携带外源 DNA 的转基因果蝇简便易行

P 因子是可转座的 DNA 片段，能引起一种杂种劣势（hybrid dysgenesis）的遗传学现象（图 22-22，框 19-4)。让我们来考虑果蝇雌性"M"株和雄性"P"株（同一种，但属不同的群体）交配的结果，其 F_1 子代常常是不育的。原因是 P 株含有多拷贝的 P 因子转座子，在来源于 M 卵的胚胎中移动，这些卵缺少抑制 P 因子移动的抑制蛋白。P 因子的切除和插入仅限于极细胞，即配子（在雄性中是精子，雌性中是卵子）的前体中。有时，P 因子插入到这些生殖细胞发育的必需基因里，就会导致交配产生的成

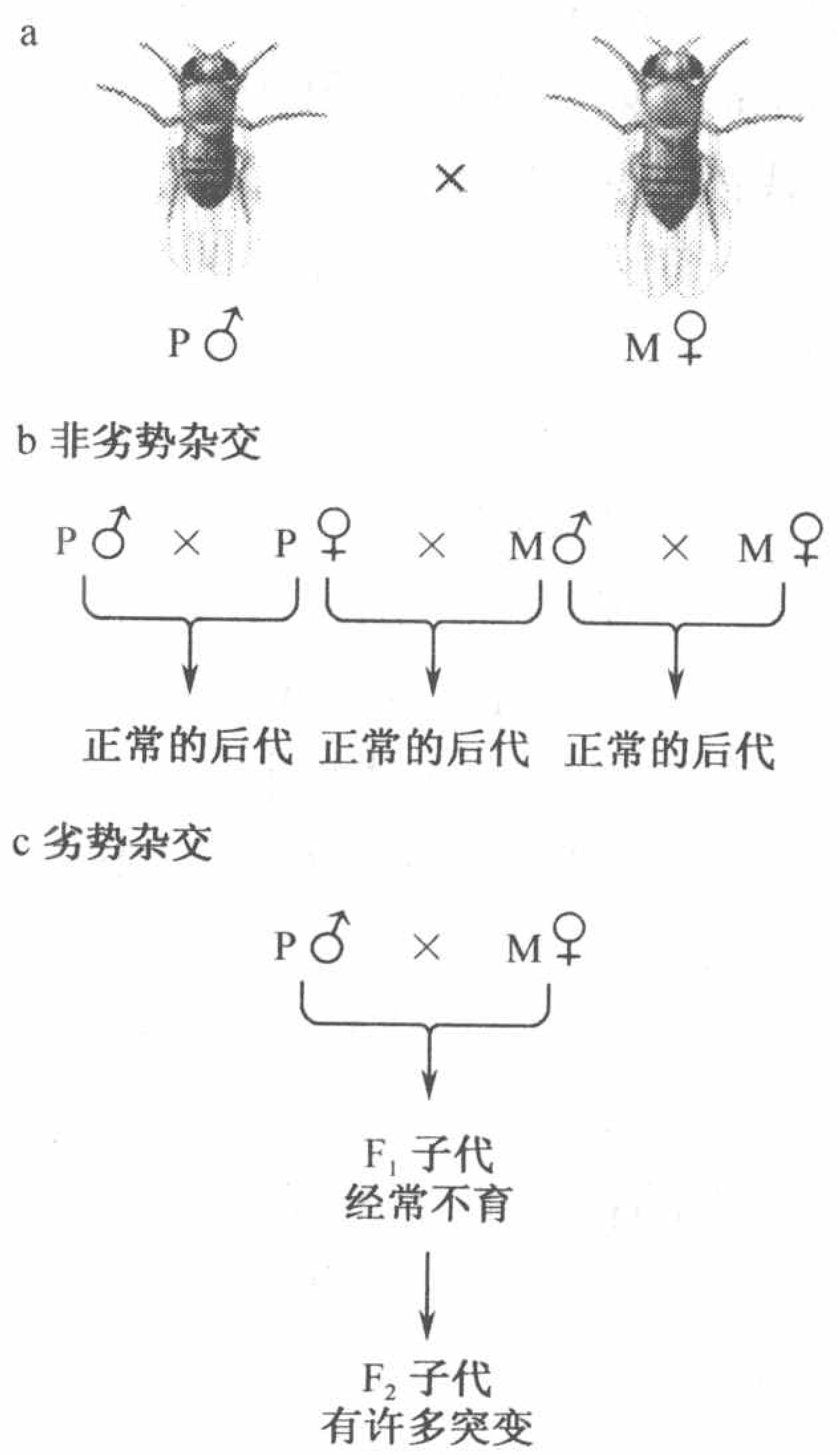


图 21-22 杂种劣势。P 因子转座子位于 P 型株中，因为阻抑物的表达，使转座子保持沉默。当 P 株与缺乏阻抑物的 M 型株交配，转座子在极细胞中移动，常常整合到生殖细胞形成所需的基因里。这就解释了这种交配后代的高频率不育。

蝇不育。

用 P 因子做载体可以将重组 DNA 引入到正常的果蝇株里（图 22-23）。全长 P 因子转座子有 3kb，它的两端含有反向重复序列，这对切除和插入是必需的。居间的 DNA 编码一个转座的阻抑物和一个促进移动的转座酶。阻抑物在发育中的 P 型株卵中表达，结果，来自 P 型株（含有 P 因子）雌性的胚胎中 P 因子不能移动，只有在那些缺乏 P 因子的雌性 M 型株所产的卵形成的胚胎中才能够移动。

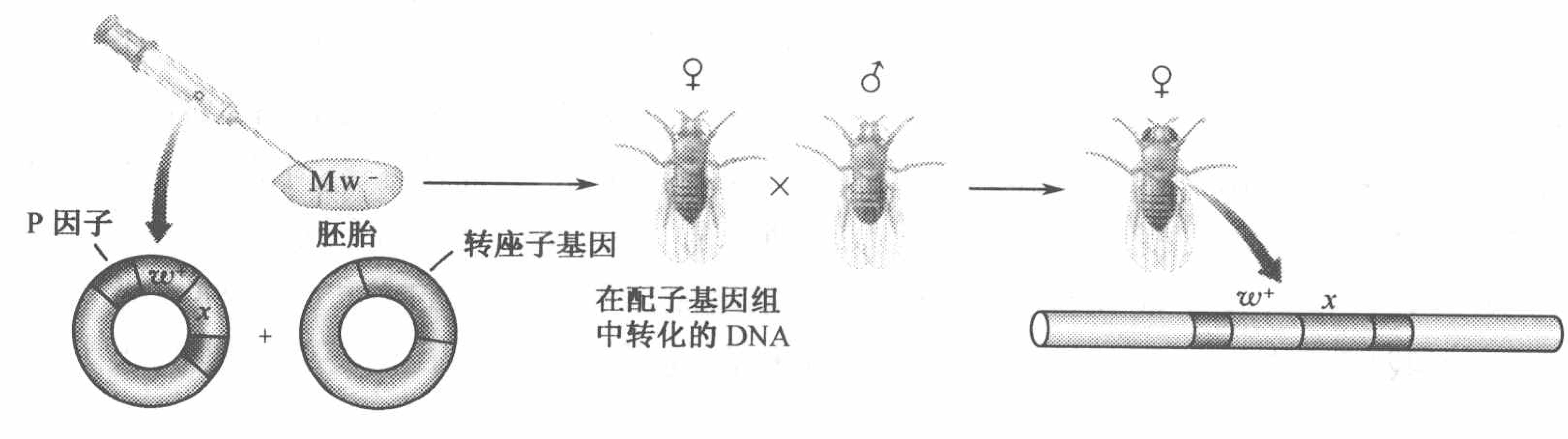


图 21-23 P 因子转化。P 因子可以用作果蝇胚胎转化的载体。所以，如同在文中所讨论的，目的序列可以插入到经过修饰的 P 因子中。这个重组分子的一个拷贝将稳定地整合到一条果蝇染色体的某一位置中。

重组 DNA 被插到编码阻抑物和转座酶基因缺乏的 P 因子中，然后 DNA 被注入早期、细胞（形成）前胚胎的后部（如同我们在第 19 章中所见，这是含有极粒的区域），同时将转座酶和重组 P 因子载体一并注入。当分裂的核进入后部，它们获得极粒、重组 P 因子 DNA 以及转座酶；极细胞从极质里出芽并脱落，重组 P 因子在极细胞里随机插入到某个位置上。不同的极细胞含有不同的插入 P 因子。重组 P 因子 DNA 和转座酶的量是经过精心试验的，所以，一个特定的极细胞平均只能有一个 P 因子的整合。胚胎发育到成体，然后与合适的试验株交配。

重组 P 因子含有一个“标记”基因如 $white^{+}$，用来注射的受体株是 $white^{-}$。测验株也是 $white^{-}$，所以所有红眼的 F_2 果蝇一定含有一个拷贝的重组 P 因子。这种 P 因子转化的方法被经常用来鉴定调控序列，如那些控制 *eve* 条带 2 表达的序列（我们在第 19 章讨论过）。此外，这种策略还被用来在各种遗传背景下检验蛋白质编码基因。

总之，就像已经提到的，果蝇提供微生物模式系统所有复杂的、经典的分子遗传学的技术。唯一明显的缺陷是缺乏通过与重组 DNA 同源重组的基因组精确操作方法，如创造基因缺失。然而，最近已经开发了这样的方法，现在正在改进以便于日常使用。有意思的是，我们将会看到，这样的操作在更复杂的小鼠生物系统中是容易进行的。不过，因为果蝇具有丰富的遗传工具，而且几十年来对它的研究使我们积累了大量的背景知识，果蝇依然是研究发育和行为的首要模式系统之一。

小鼠

如果以秀丽新小杆线虫和果蝇为标准，小鼠的生活周期显得缓慢并且操作难度更大。胚胎发育或妊娠需要 3 个星期，新生的小鼠还要过 5～6 个星期才能达到青春期。于是，小鼠的正常生活周期大约是 8～9 个星期，比果蝇的长 5 倍多。然而，由于它在进化树上的位置，鼠类享有一个特殊的待遇：它是哺乳动物，因此和人类有亲缘关系。当然，黑猩猩和其他更高级的灵长类与人类之间有更近的亲缘关系，但我们不容易用它们进行各种在小鼠中能够进行的实验。

于是，小鼠的研究使我们能够将低等生物如线虫和果蝇研究中发现的基本原理与人类疾病联系起来。例如，果蝇的 *patched* 基因编码了 Hedgehog 受体的重要组成元件（第 19 章）。果蝇中缺乏野生 *patched* 基因活性的突变体，其胚胎会表现出各种特定的身体结构缺陷，小鼠中的同源基因在发育中也是重要的。然而出乎意料的是，在鼠和人类中，有些 *patched* 突变体导致癌症，如皮肤癌，在果蝇中的分析没能揭示这样的功能。另外，已经开发了能在正常的小鼠中有效地敲除特定基因的方法，这种“敲除”技术影响了我们对人类发育、行为和疾病基本机制的了解。下面我们将简短地回顾小鼠作为实验系统的显著特征。

小鼠的染色体组成与人类相似：小鼠有 19 条常染色体（人类有 22 条），还有 X 和 Y 性染色体。在小鼠和人类之间有着大量的同线性：小鼠染色体的许多区域有着与人类染色体的“同源”区域相应的基因（以同样的顺序）。小鼠基因组已经被测序和组装。如

在第 19 章中所讨论的，小鼠与人类的基因组几乎有相同的基因组成：两者都含有约 30 000 个基因，其中 85%以上的基因是一一对应的；小鼠与人类基因组的差别大多数，尽管不是全部，源于两个谱系中的某些基因家族选择性重复的不同。比较基因组分析证实了我们早已知道的事情，即小鼠是研究人类发育和疾病的优良模型。

鼠的胚胎发育依赖干细胞

小鼠的卵很小，难以操作。与人类的卵一样，它们的直径只有 100μm，这使像在斑马鱼和青蛙中进行的那种嫁接移植难以在小鼠中进行，但把重组 DNA 引入小鼠的细胞系从而产生转基因小鼠的显微注射法已经建立，下文将会对此进行讨论。另外，即使在发育的较早时期，也可能获得足够的小鼠胚胎，进行原位杂交以及特定基因表达模式的直观检测。这种直观检测方法适用于正常的胚胎和在特定的遗传位点发生缺陷的突变体。

图 22-24 显示小鼠胚胎发生的概况。早期小鼠胚胎的初始分裂非常缓慢，仅以平均每 12～24h 一次的频率发生。在 16 细胞期可以观察到最早的明显的细胞类型分化，这时的胚胎叫桑椹胚（morula）（图 22-24，第 6 幅图）。外周细胞形成的组织与胚胎形成无关，但能发育成胎盘，内部细胞则产生内细胞团（ICM）。在 64 细胞期，只有 13 个 ICM 细胞，这些细胞浆形成成鼠的所有组织。ICM 是胚胎干细胞的主要来源，可以在培养基里培养，并且，一旦加入合适的生长因子，就可被诱导形成各种类型的成体细胞。人类干细胞已经成为相当规模的社会争论的话题，但确实提供了一个可更新的组织来源，以替代很多变性疾病，如糖尿病和阿尔茨海默氏症中的缺陷细胞。

在 64 细胞期（大约受精后 3～4d），小鼠胚胎，此时称为胚泡（blastocyst），最终准备着床。胚泡和子宫壁的相互作用导致了胎盘的形成，胎盘是除了原始的产蛋的鸭嘴兽之外所有哺乳动物都具有的特征。胎盘形成后，胚胎进入原肠期，由此，ICM 形成所有 3 个胚层：内胚层、中胚层

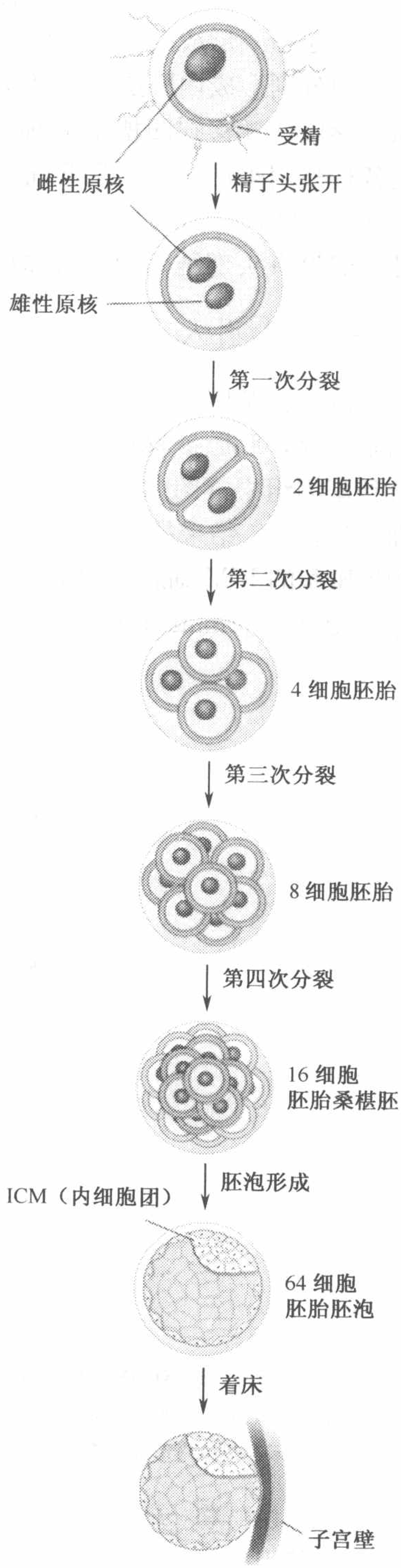


图 22-24　小鼠胚胎发生概观。

和外胚层。之后不久，具有脑、脊髓和内部器官如心脏和肝脏的胎儿就形成了。

小鼠原肠胚形成的第一期中 ICM 分成了两个细胞层：里面的下胚层和外面的上胚层，它们分别形成内胚层和外胚层。一个叫原条（primitive streak）的凹槽沿着上胚层的纵向形成，迁移到凹槽里的细胞形成了里面的中胚层。原条的前端称为节（node）；它是各种被用来形成胚胎贯穿前后的中轴信号分子的来源，这些信号分子包括两个分泌型的 TGF-β 信号传导的抑制物，Chordin 和 Noggin。缺失这两个基因的双突变小鼠胚胎发育成没有头部结构如前脑和鼻子的胎儿。

把外源 DNA 引入小鼠胚胎并不复杂

用以在转基因小鼠中有效表达重组 DNA 的显微注射法已经建立。DNA 被注射进卵细胞原核，胚胎被移置入一个雌鼠的输卵管，使其着床并发育，注射的 DNA 随机地整合到基因组中（图 22-25）。整合的效率相当高，并通常发生在发育的早期，往往是在单细胞胚胎期。结果是融合基因插入大多数或所有的胚胎细胞，形成包括成鼠的体细胞组织和生殖系的 ICM 细胞。用这种简单的显微注射方法产生的转基因小鼠大约 50%表现出生殖系转化（germline transformation），也就是说，它们的后代也含有外源重组 DNA。

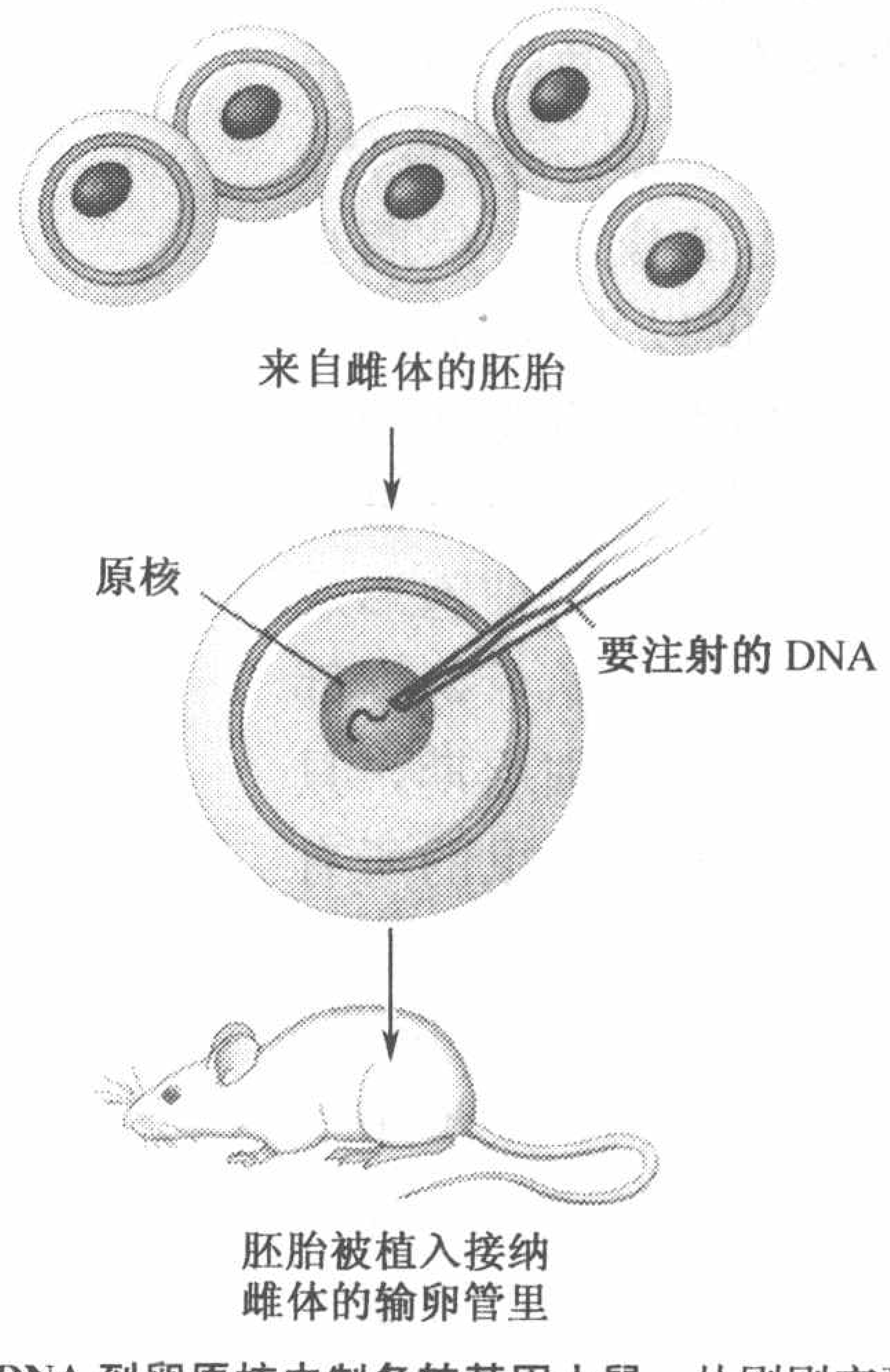


图 22-25 通过显微注射 DNA 到卵原核内制备转基因小鼠。从刚刚交配的雌鼠中取出单细胞胚胎。重组 DNA 被注射到细胞核里，然后，胚胎被植入一个假孕母鼠的输卵管内。几天后，胚胎着床，并最终形成含有重组 DNA 的整合拷贝的胎儿。

举例来说，将 *Hoxb-2* 基因的增强子和 *lacZ* 报道基因连接形成融合基因。从携带这个报道基因的转基因株中获得胚胎和胎儿，并染色以观察 *lacZ* 的表达模式。在此例中，观察到显色区在后脑中（图 21-26），转基因小鼠已经被用来研究好多个调节序列，包括那些调节 β-珠蛋白基因和 *HoxD* 基因的序列。这两个复杂的位点都含有远程调控元件（分别

是 LCR 和 GCR），它们能协调相距几十万碱基的不同基因的表达（第 17 章）。

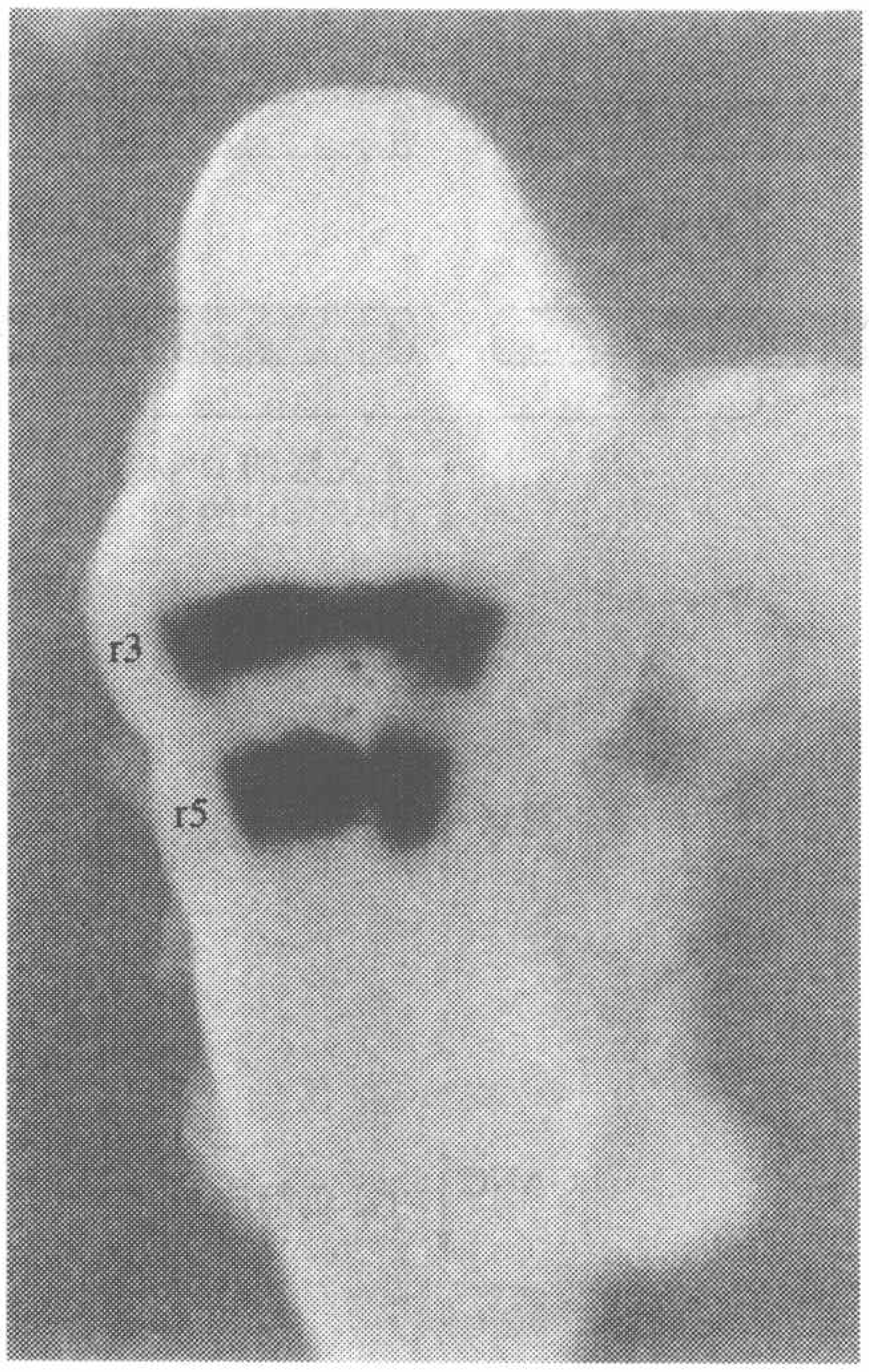


图 22-26 转基因小鼠胚胎的原位表达模式。这是由部分 *Hoxb-2* 基因的调节区域与 *lacZ* 报道基因连接在一起获得的转基因小鼠。从转基因雌鼠取出胚胎，并染色以显示 β-半乳糖苷酶（lacZ）活性的位置。10.5d 的胚胎后脑区域有两个主要的染色条带。图中胚胎头朝上尾向下。（来源：Nonchev et al. 1996. *Proc. Natl. Acad. Sci.* 93：9339-9345，Fig. 1c.）

同源重组为单个基因的选择性去除提供了机会

小鼠转基因的一个强有力的方法是中断，或“敲除”单个遗传位点的能力，这使人类疾病的小鼠模型的制备成为可能。例如，*p53* 基因编码一个调节蛋白，它激活 DNA 修复所需的基因的表达，已经发现它与许多人类癌症有关。*p53* 功能丧失时，由于 DNA 突变的快速积累，癌细胞就具有更高的侵染性。已经建立了一个小鼠株，它除了 *p53* 基因被敲除外，其余完全正常，这些小鼠非常容易患癌症，死得很早。但这些小鼠有望用于检测潜在的供人类使用的药物和抗癌剂。尽管果蝇中也有 *p53* 基因，也分离了突变体，但它并不能像小鼠模型那样用于药物发现。

基因打断实验是用胚胎干细胞（ES）来做的（图 22-27），这些细胞是通过培养小鼠胚泡而获得，ICM 细胞只增殖但不分化。为此构建一个含有突变目的基因的重组 DNA。例如，在基因的起始端附近删除一小段区域，而改变一个特定目标基因的蛋白质编码区，从编码的蛋白质中删除必需氨基酸的密码子，并在剩余的编码序列中产生移码。经过改变的靶基因和耐药基因（如具有对新霉素抗性的 *NEO* 基因）相连接，只有那些含有转基因的 ES 细胞可以在含有抗生素的培养基里生长。*NEO* 基因被置于经过

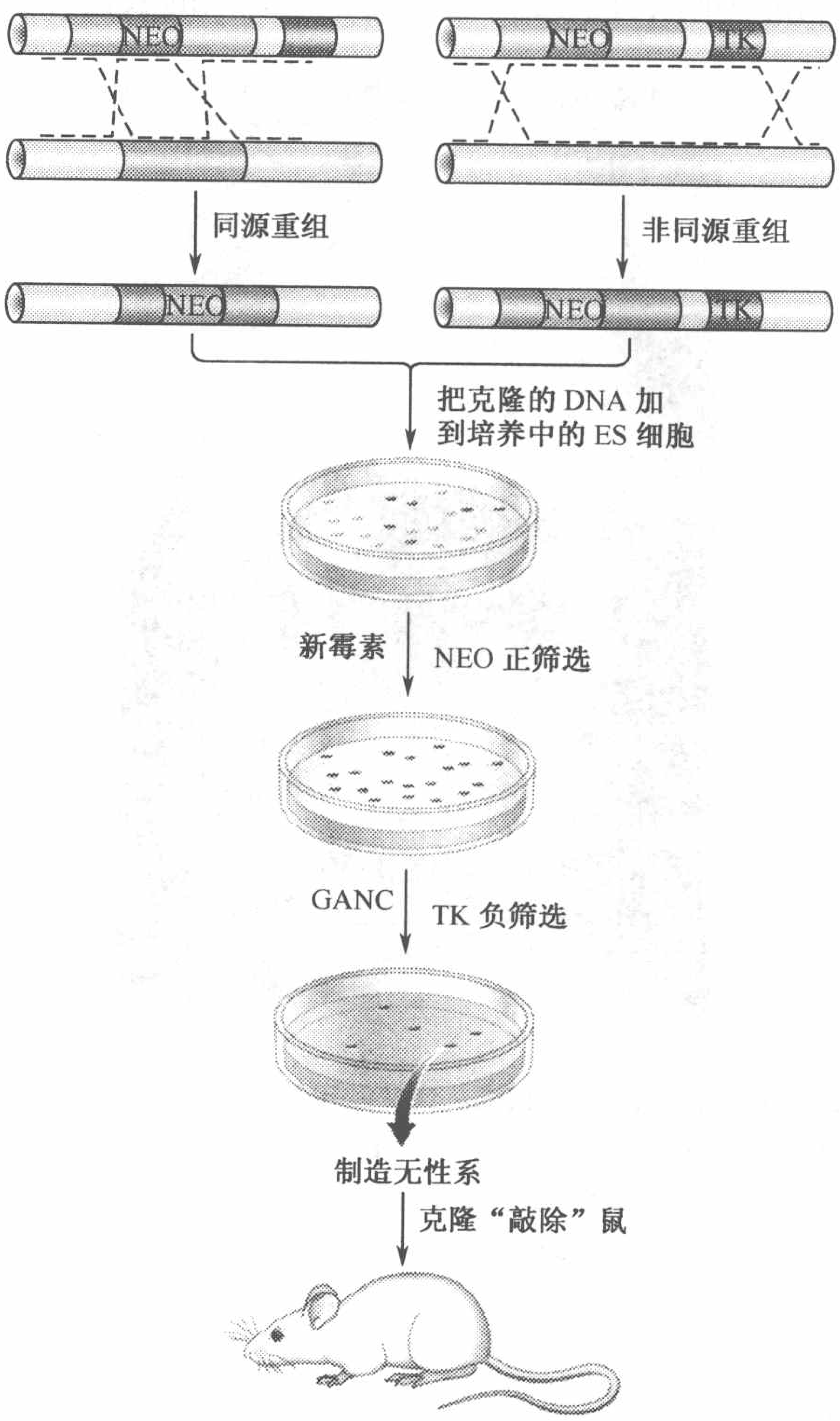


图 22-27 通过同源重组进行基因敲除。本图概括了制备任意基因缺失的细胞株的方法。目标基因（以绿色显示）内发生的同源重组导致了 *NEO* 基因的插入并破坏了那个基因。非同源重组或随机重组可以导致含有 *NEO* 基因的中断基因和胸苷激酶（TK）的基因的插入。携带这两个转基因片段的克隆在新霉素中都能存活，但同时携带 TK 的克隆随后在丙氧鸟苷（gancyclovir，GANC）中经过反选择。于是，只有含有携带 *NEO* 基因和靶基因的转基因片段的克隆才能存活。一旦产生突变细胞就可以被克隆并培养成敲除该基因的小鼠（图 22-25）。

改变的靶基因的下游，但位于与染色体同源的区段侧翼区域上游，这样转基因片段和染色体的双重重组将导致靶基因被突变基因和抗药基因所取代（或者用另一种办法，*NEO* 基因可以插入目标基因）。

然而，也存在高发性的非同源重组，即发生在同源的体内基因之外的不正确重组。为了富集同源重组，重组载体还包含了一个标记基因——胸苷激酶基因（TK），可以使用药物丙氧鸟苷进行负筛选，丙氧鸟苷在激酶作用下将转化为一个有毒的化合物。在载体上，胸苷激酶基因位于与染色体同源的区域之外。通过同源重组使突变基因整合到染

色体的转化子会失去胸苷激酶基因，但通过非同源重组整合入染色体的转化子含有胸苷激酶的完整载体，于是就会被筛除掉。

通过上述过程可以获得重组 ES 细胞，其中，一份目标基因与突变的等位基因相对应。收集这些重组 ES 细胞并注射到正常胚泡的 ICM 里。杂种胚胎植入宿主鼠的输卵管，并让它们发育到相应时期。有些源于杂种胚胎的成体拥有转化了的生殖系，于是产生含有突变目标基因的单倍体配子。最初用作转化的和同源重组检测的 ES 细胞产生体细胞组织和生殖系。一旦获得了具有转化生殖细胞的小鼠，就将其与同胞小鼠进行交配，以获得纯合突变体。有时，由于这些突变体是致死的，必须在胚胎时就对它们进行分析。而其他非致死基因，突变体胚胎能发育成生长完全的小鼠，然后用各种技术去检测。

小鼠中存在 epi-遗传

遗传操作的胚胎研究发现了一个奇怪的非孟德尔遗传现象，或者表观遗传现象，这个现象还称为亲本印记（parental imprinting）（图 22-28）。基本的概念是某些基因的两

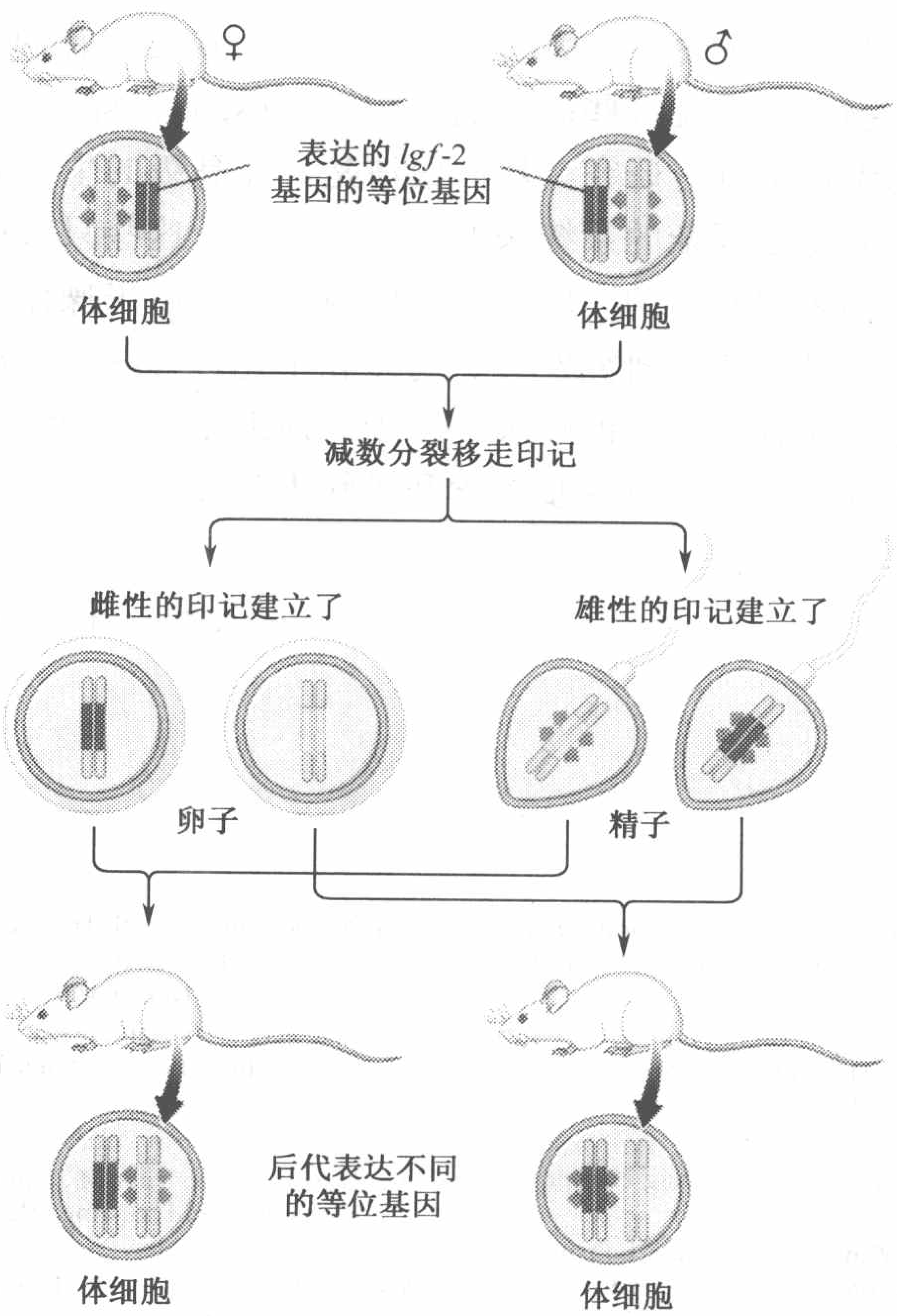


图 22-28　小鼠中的印记。小鼠中特定基因的一个等位基因的永久沉默。如文中所略述和在第 17 章中所详细描述的，印记保证了小鼠的 *Igf-2* 基因在每个细胞中只有一个拷贝是表达的。被表达的总是那个源于父本染色体的拷贝。

个等位基因中只有一个是有活性的，这是因为另一个拷贝在发育中的精子细胞或发育中的卵子细胞里被选择性地失活了。让我们举个 *lgf-2* 基因的例子，它编码一个类似胰岛素的生长因子，在发育胎儿的内脏和肝脏中表达。只有从父本那里继承下来的*lgf-2*等位基因在胚胎里活跃地表达；另一个拷贝，尽管在序列上完全正常，却是没有活性的。*lgf-2* 基因的母本和父本拷贝在活性上的这种差别源自于一个与之关联的抑制 *lgf-2* 表达的沉默子 DNA 的甲基化。在精子发育过程中，DNA 是甲基化的，所以 *lgf-2* 基因能够在发育的胚胎中活化，甲基化抑制了沉默子；相反，在发育中的卵子里这个沉默子 DNA 不被甲基化。于是，从母本继承来的 *lgf-2* 等位基因是沉默的。也就是说，对未来在胚胎中的表达，这个基因的父本拷贝做了“印记”，在这里，就是被甲基化了。这个例子在第 17 章里有详细的讨论。

在鼠和人类中大约有 30 个被印记的基因。其中许多控制发育中胎儿的生长，包括上述例子中的 *lgf-2*。有人提出印记是为了保护母亲免遭她的胎儿对其伤害而进化来的。*lgf-2* 蛋白促进胎儿的生长。母亲试图通过失活该基因的母本拷贝来限制这种生长。

在本书中我们已经讨论了每一个生物必须怎样保持和复制它的 DNA 来生存、适应和繁殖。在大多数生物中，实现这些基本的生物学目标的总体战略是相似的，因此可以相当成功地用简单的生物来进行研究。然而很清楚，在高级的生物中发现的更复杂的过程，如分化和发育，需要更复杂的系统来调节基因表达，这些只有用更复杂的生物才能研究。我们已经看到很多强有力的实验技术可以被成功地用来操作小鼠，并探索各种复杂的生物学问题。小鼠已经成为研究发育、遗传和生化过程的一个出色的模式系统，这些过程很可能发生在进化程度更高的哺乳动物中。最近小鼠基因组序列的公布及其注释更加强调了小鼠作为模式生物在进一步探索和了解人类发育和疾病研究中的重要性。

（吴琳 译 刘国振 朱金桂 校）

参考文献

Burke D., Dawson D., and Stearns T. 2000. *Methods in yeast genetics.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Hartwell L.H., Hood L., Goldberg M.L., Reynolds A.E., Silver L.S., and Veres R.C. 2004. *Genetics: From genes to genomes*, 2nd ed. McGraw–Hill, New York.

Miller J.H. 1972. *Experiments in molecular genetics.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Nagy A., Gertsenstein M., Vintersten K., and Behringer R. 2003. *Manipulating the mouse embryo*, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sambrook J. and Russell D.W. 2001. *Molecular cloning: A laboratory manual*, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Snustad D.P. and Simmons M.J. 2002. *Principles of genetics*, 3rd ed. John Wiley and Sons, New York.

Stent G.S. and Calendar R. 1978. *Molecular genetics: An introductory narrative.* W.H. Freeman and Co., San Francisco.

Sullivan W., Ashburner M., and Hawley R.S. 2000. Drosophila *protocols.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Wolpert L., Beddington R., Lawrence P., Meyerowitz E., Smith J., and Jessell T.M. 2002. *Principles of development*, 2nd ed. Oxford University Press, Oxford.

索引

P

Q

R

S

T

［Ｇｅｎｅｒａｌ　Ｉｎｆｏｒｍａｔｉｏｎ］
书名＝基因的分子生物学
作者＝（美）Ｊ．Ｄ．Ｉ沃森，Ｔ．Ａ．贝克，Ｓ．Ｐ．贝尔等著；杨焕明等译
页数＝８２４
ＳＳ号＝１２３８６２３７
出版日期＝２００９．０７

ＤＮＡ能够形成规则的螺旋
ＲＮＡ形成多种多样的结构
蛋白质构成单元的化学特性
肽键
蛋白质具有４级结构
框５－１　蛋白质结构的确定
α螺旋和β折叠是常见的二级结构形式
氢键的排列方式决定蛋白质的特定构型
α螺旋相互结合形成卷曲螺旋
大多数蛋白质由两个或三个结构域装配而成
框５－２　大蛋白质通常由数个小多肽链组成
蛋白质由少数几种结构模式所组成
蛋白质的不同功能来自于结构域组合的多样性
弱化学键决定蛋白质在ＤＮＡ和ＲＮＡ分子上的正确位置
蛋白质沿ＤＮＡ扫描寻找特定的ＤＮＡ结合位点
蛋白质识别ＲＮＡ的不同策略
变构：通过改变蛋白质的外形调节其功能
通过小配体、蛋白质－蛋白质相互作用和蛋白质修饰等例子了解变构调控的结构基础
并非所有的蛋白质调控都由变构介导
小结
参考文献
第２篇　基因组的维持第６章　ＤＮＡ和ＲＮＡ的结构
ＤＮＡ的结构
ＤＮＡ由多核苷酸链构成
每个碱基都有它首选的异构体
双螺旋的两条链通过碱基反向平行配对连接在一起
双螺旋的两条链序列互补
氢键对碱基配对特异性的重要作用
碱基会从双螺旋中翻转出来
ＤＮＡ通常是右手双螺旋
双螺旋有大沟和小沟
框６－１　云母实验表明溶液中双螺旋ＤＮＡ每一螺周含有１０．５个碱基对
大沟中有丰富的化学信息
双螺旋以多重构象存在
重要实验
框６－２　如何通过Ｘ射线胶片中的斑点揭示ＤＮＡ结构
ＤＮＡ有时可形成左手螺旋
ＤＮＡ双链可以分开（变性）和复性
一些ＤＮＡ为环状分子
ＤＮＡ拓扑学
连环数是共价闭环ＤＮＡ的固有拓扑特性
连环数由扭转数和缠绕数共同决定
Ｌｋ°是生理条件下完全松弛态ｃｃｃＤＮＡ的连环数
细胞中ＤＮＡ呈负超螺旋
真核细胞中核小体引入了负超螺旋
拓扑异构酶可使超螺旋ＤＮＡ解旋
原核生物中存在引导ＤＮＡ超螺旋形成的特殊拓扑异构酶
拓扑异构酶也可以解链和松弛ＤＮＡ分子
拓扑异构酶通过蛋白质－ＤＮＡ的共价连接裂解ＤＮＡ链或使它们重新连接
拓扑异构酶构成“酶桥”让ＤＮＡ片段往来穿梭
ＤＮＡ拓扑异构体可被电泳分离
框６－３　ＤＮＡ环的拓扑学特性证明了ＤＮＡ每周螺旋含有１０．５个碱基对的螺旋周期性
溴乙锭离子可使ＤＮＡ解旋
ＲＮＡ的结构
ＲＮＡ含有核糖和尿嘧啶，通常是单链
ＲＮＡ链自身折叠形成局部双螺旋，类似Ａ－ＤＮＡ
ＲＮＡ可折叠成复杂的三级结构
一些ＲＮＡ可以是酶类

转座
　　一些遗传因子通过转座被移位到染色体的新位置上
三类基本的转座因子
　　ＤＮＡ转座子上带有一个转座酶基因，两端是重组位点
　　转座子包括自主转座子和非自主转座子
　　类病毒反转录转座子和反转录病毒带有末端重复序列和两个对重组很重要的基因
　　像基因的聚腺苷酸反转录转座子
　　ＤＮＡ转座的剪切－粘贴机制
　　剪切－粘贴转座中的中间体由缺口修复来完成
　　ＤＮＡ转座中断裂非转移链的多种机制
　　ＤＮＡ的复制性转座
　　类病毒反转录转座子和反转录病毒通过一个ＲＮＡ中间体实现移位
　　框１１－３　　反转录病毒ｃＤＮＡ的合成途径
　ＤＮＡ转座酶和反转录整合酶同属于一个蛋白超家族
　　聚腺苷酸（ｐｏｌｙ－Ａ）反转录转座子通过一种“反向剪切”机制进行移位
转座子及其调控的实例
　　框１１－４　　玉米转座因子及转座子的发现
　　ＩＳ４家族的紧密型转座因子通过多种机制实现对拷贝数的控制
　　Ｔｎ１０的转座伴随着细胞ＤＮＡ的复制
　　框１１－５　　转座的靶免疫机制
　　Ｍｕ噬菌体是一种异常活跃的转座子
　　Ｍｕ噬菌体通过靶位点免疫来避免其转座到自身ＤＮＡ上
　　Ｔｃｌ／ｍａｒｉｎｅｒ因子是真核细胞中极其成功的ＤＮＡ转座因子
　　酵母Ｔｙ因子转座到基因组的安全港
　　ＬＩＮＥ因子能够推动自身转座，还能转座细胞ＲＮＡ
Ｖ（Ｄ）Ｊ重组
　　Ｖ（Ｄ）Ｊ重组中的早期活动是通过一种类似于转座子切除的机制实现的
小结
参考文献
第３篇　　基因组的表达第１２章　　转录机制
ＲＮＡ聚合酶和转录周期
　　ＲＮＡ聚合酶具有不同形式，但有许多共同特性
　　由ＲＮＡ聚合酶进行的转录是由一系列步骤完成的
　　转录起始包括三个定义明确的步骤
细菌的转录周期
　　细菌的启动子在强度和序列上是多样的，但是具有某些明确的特征
框１２－１　　共有序列
　σ因子介导聚合酶与启动子的结合
向开放式复合体的转变过程涉及ＲＮＡ聚合酶和启动子ＤＮＡ的结构变化
　　转录由ＲＮＡ聚合酶起始，不需要引物
　　转录起始阶段，ＲＮＡ聚合酶保持位置不变并将下游ＤＮＡ拉向自己
　　启动子逃离涉及聚合酶－启动子相互作用及聚合酶核心－σ相互作用的破坏
延伸聚合酶是一个边合成边校对ＲＮＡ的向前行进的机器
框１２－２　　单亚基ＲＮＡ聚合酶
　ＲＮＡ聚合酶可能变得停滞不前，需要挪走
　　转录由ＲＮＡ序列内部信号终止
真核生物的转录
　　ＲＮＡ聚合酶Ⅱ的核心启动子由４个不同序列的元件组合而成
　　ＲＮＡ聚合酶Ⅱ在启动子上和通用转录因子形成前起始复合体
　　启动子逃离需要聚合酶“尾巴”的磷酸化
　　ＴＢＰ用一个插入双螺旋小沟的β折叠来结合并扭曲ＤＮＡ
　　其他通用转录因子在起始中也有特殊作用
　　体内的转录起始需要其他蛋白质参与，包括中介蛋白复合体
　　中介蛋白包含很多亚基，其中有些亚基从酵母到人类是保守的
　　一组新的因子激发ＰｏｌⅡ延伸和ＲＮＡ校对
　　延伸ＲＮＡ聚合酶在行进过程中必须应对组蛋白
　　延伸聚合酶与各种ＲＮＡ加工所需要的一组新的蛋白质因子相互作用
由ＲＮＡ聚合酶Ⅰ和Ⅲ催化的转录

ＲＮＡ聚合酶Ⅰ和Ⅲ识别不同的启动子，使用不同组别的转录因子，但还是需要ＴＢＰ
　　ＰｏｌⅢ启动子位于转录起始位点的下游
小结
参考文献
第１３章　　ＲＮＡ剪接
ＲＮＡ剪接的化学基础
　　ＲＮＡ序列决定剪接位点
　　框１３－１　　腺病毒和ＲＮＡ剪接的发现
　　当内含子两边的外显子连接时，内含子是以套索结构被除去的
　　不同的ＲＮＡ分子的外显子可以通过反式剪接的方式连接起来
剪接体
　ＲＮＡ剪接是由一个叫做剪接体的大复合体执行的
剪接过程
　　剪接体内的组装、重排与催化反应：剪接过程
　　自剪接内含子的存在表明ＲＮＡ可以催化ＲＮＡ剪接
　　　Ⅰ类自剪接内含子释放出线形内含子，而不是一个套索结构
　　框１３－２　　Ⅰ类自剪接内含子转化为核酶
　　剪接体怎样可靠地确定剪接位点？
　　少数内含子是由另一组ｓｎＲＮＰ组成的另一种剪接体来剪接的
替换剪接
　　通过替换剪接一个基因可以得到多个产物
确保互不相容性剪接的几种机制
关于果蝇的Ｄｓｃａｍ基因的特例：大范围的互不相容性剪接
　Ｄｓｃａｍ外显子６的互不相容性剪接无法用任何一种标准机制来解释，而要用一种新的策略来解释
可变剪接受激活因子和抑制因子调控
可变剪接的调控决定了果蝇的性别
框１３－３　　停泊位点和选择子序列的鉴定
框１３－４　　Ｐｒｅ－ｍＲＮＡ剪接错误导致人类疾病
外显子改组（Ｅｘｏｎ　Ｓｈｕｆｆｌｉｎｇ）
　　外显子通过重组的方式完成改组，从而产生了编码新蛋白质的基因
ＲＮＡ编辑
　　ＲＮＡ编辑是改变ｍＲＮＡ序列的另一种方法
　　引导ＲＮＡ指导尿嘧啶的插入和删减
　　框１３－５　　脱氨酶和ＨＩＶ
ｍＲＮＡ转运
　　一旦完成加工，ｍＲＮＡ会包装好并从细胞核输出到细胞质用于翻译
小结
参考文献
第１４章　　翻译
信使ＲＮＡ
　　多肽链是由可读框规定的
　　原核细胞ｍＲＮＡ具有核糖体结合位点，用以募集翻译机器
　　真核细胞ｍＲＮＡ在５′和３′端被修饰，以促进翻译
转移ＲＮＡ
　　ｔＲＮＡ是密码子和氨基酸之间的转配器
　　框１４－１　　ＣＣＡ添加（ＣＣＡ－Ａｄｄｉｎｇ）酶：无模板合成ＲＮＡ
　　ｔＲＮＡ都有三叶草型的二级结构
　　ｔＲＮＡ具有Ｌ状的三维结构
连接氨基酸到ｔＲＮＡ上
　　氨基酸通过高能酰基连接到ｔＲＮＡ　３′端的腺苷酸上使ｔＲＮＡ负载
　　氨基酰ｔＲＮＡ合成酶分两步使ｔＲＮＡ负载
　　每种氨基酰ｔＲＮＡ合成酶连接一个氨基酸到一个或多个ｔＲＮＡ上
　　ｔＲＮＡ合成酶识别同族ｔＲＮＡ（ｃｏｇｎａｔｅ　ｔＲ－ＮＡ）的独特结构
　　氨基酰ｔＲＮＡ的形成是非常精确的
　　某些氨基酰ｔＲＮＡ合成酶利用编辑口袋来高精度地负载ｔＲＮＡ
　　核糖体不能辨别ｔＲＮＡ的负载是否正确
　　框１４－２　　硒化半胱氨酸（ｓｅｌｅｎｏｃｙｓｔｅｉｎｅ）
核糖体

核糖体是由一个大亚基和一个小亚基组成的
大、小亚基在翻译的每次循环过程中都经历结合与分离
新的氨基酸连接在延伸肽链的C端
肽键是通过将正在延伸的多肽链从一个ｔＲＮＡ转移到另一个ｔＲＮＡ而形成的
核糖体ＲＮＡ是核糖体中的结构和催化决定因子
核糖体有３个ｔＲＮＡ结合位点
穿过核糖体的通道可以使ｍＲＮＡ和多肽链进出核糖体
翻译的起始
原核细胞的ｍＲＮＡ最初是通过与ｒＲＮＡ的碱基配对而募集到小亚基上
负载着修饰甲硫氨酸的特定的ｔＲＮＡ直接结合到原核细胞的小亚基上
三种起始因子指导包含ｍＲＮＡ和起始ｔＲＮＡ的起始复合物的装配
真核细胞的ｍＲＮＡ通过５′帽吸引核糖体
框１４－３　ｕＯＲＦ和ＩＲＥＳ：证明规则的特例
起始密码子是从ｍＲＮＡ的５′端向下游“扫描发现”的
翻译起始因子使真核ｍＲＮＡ保持环状
翻译延伸
氨基酰ｔＲＮＡ由延伸因子ＥＦ－Ｔｕ送达Ａ位点
核糖体利用多种选择机制排斥错误配对的氨基酰ｔＲＮＡ
核糖体是核酶
肽键形成和延伸因子ＥＦ－Ｇ促进了ｔＲＮＡ和ｍＲＮＡ的易位
ＥＦ－Ｇ通过取代结合在Ａ位点的ｔＲＮＡ来驱动易位
ＥＦＴｕ－ＧＤＰ和ＥＦ－ＧＧＤＰ在参加新一轮延伸之前必须将ＧＤＰ换成ＧＴＰ
形成肽键的一个循环要消耗两个ＧＴＰ分子和一个ＡＴＰ分子
框１４－４　ＧＴＰ结合蛋白、构象转变以及翻译事件的忠实性和次序
翻译终止
释放因子在终止密码子的作用下终止翻译
Ｉ类释放因子的一小段区域识别终止密码子并催化多肽链的释放
ＧＤＰ／ＧＴＰ交换和ＧＴＰ水解调控Ⅱ类释放因子的功能
核糖体回收因子模仿ｔＲＮＡ
框１４－５　抗生素通过阻遏翻译的某些特定步骤抑制细胞分裂
翻译的调控
蛋白质或ＲＮＡ在核糖体结合位点附近结合下调细菌翻译的起始
原核翻译调控：核糖体蛋白是其自身合成的翻译抑制因子
真核翻译总体调节因子以ｍＲＮＡ识别及起始ｔＲＮＡ核糖体结合所需的主要因子为靶标
通过对ｍＲＮＡ特异的４Ｅ－ＢＰｓ进行的翻译空间调控
一个铁离子调控的ＲＮＡ结合蛋白控制转铁蛋白的翻译
酵母转录激活因子Ｇｃｎ４的翻译受到上游短ＯＲＦ及三元复合物丰度的调控
依赖翻译的ｍＲＮＡ和蛋白质的稳定性调节
ＳｓｒＡ　ＲＮＡ援救翻译缺陷ｍＲＮＡ的核糖体
真核细胞降解不完整的或终止密码子提前的　ｍＲＮＡ
小结
参考文献
第１５章遗传密码
遗传密　码具简并性
认识密码子组成的规律
反密码子的摆动配对
三个密码子指导链的终止
遗传密码是如何破译的？
通过人工合成的ｍＲＮＡ激发氨基酸的掺入
ｐｏｌｙ－Ｕ编码多聚Ｐｈｅ
混合多聚核苷酸使其他密码子的组成得以确定
ｔＲＮＡ和特定的三联密码子结合
重复共聚物确定密码子
遗传密码遵循的三条规律
改变遗传密码的三种点突变
遗传密码是以３个碱基为阅读单位的遗传学证据
抑制突变可以存在于相同或不同的基因中
基因间抑制涉及突变ｔＲＮＡ

无义突变抑制也能识别正常的终止信号
证实遗传密码的正确性
遗传密码几乎通用
小结
参考文献
第４篇　　调　　　　控第１６章　　原核生物的基因调控
转录调控的原理
基因表达由调控蛋白控制
很多启动子通过协助ＲＮＡ聚合酶结合ＤＮＡ的活化子和阻碍两者结合的抑制子进行调控
某些活化子通过变构和调控ＲＮＡ聚合酶对结合后的转录步骤起作用
远程激活和ＤＮＡ环化
协同结合和变构在基因调控中有多种作用
抗终止作用及其延续：基因调控并不局限于转录起始调控
转录起始的调控：原核生物的实例
活化子和抑制子共同控制ｌａｃ基因
ＣＡＰ和Ｌａｃ抑制子对ＲＮＡ聚合酶结合ｌａｃ启动子的作用是相反的
ＣＡＰ具有独立的激活表面和ＤＮＡ结合表面
框１６－１　　活化子旁路实验
ＣＡＰ和Ｌａｃ抑制子以相同的结构模体结合ＤＮＡ
Ｌａｃ抑制子和ＣＡＰ的活性由其信号变构控制
框１６－２　　Ｊａｃｏｂ、Ｍｏｎｏｄ及其基因调控的观点
组合控制：ＣＡＰ也控制其他基因
选择性的σ因子指导ＲＮＡ聚合酶选择性地结合启动子
ＮｔｒＣ和ＭｅｒＲ：以变构而非募集作用的转录活化子
ＮｔｒＣ具有ＡＴＰａｓｅ活性且其作用位点距基因较远
ＭｅｒＲ通过扭曲ＤＮＡ激活转录
某些抑制子在启动子上滞留而非排斥ＲＮＡ聚合酶
ＡｒａＣ和抗激活作用对ａｒａＢＡＤ操纵子的控制
λ噬菌体：调控的层次
基因表达的选择性模式控制裂解和溶原性生长
调控蛋白及其结合位点
λ抑制子与操作子位点协同结合
框１６－３　　浓度、亲和力和协同结合
抑制子和Ｃｒｏ以不同模式结合控制裂解和溶原生长
溶原性诱导需要λ抑制子的蛋白水解切割
抑制子的负自我调控需远程相互作用和大的ＤＮＡ环
框１６－４　　λ转换的进化
λｃⅡ，一种控制溶原或裂解、生长或感染新宿主的活化子
侵染特定细菌的噬菌体颗粒的数目影响侵染走向溶菌或溶原
框１６－５　　遗传学方法确定参与裂解／溶原选择的基因
Ｅ．ｃｏｌｉ的生长条件控制ＣⅡ的稳定性并因而控制裂解／溶原的选择
λ发育中的转录抗终止作用
逆回调控：ＲＮＡ合成和稳定性控制相互影响并决定基因表达
小结
参考文献
第１７章　　真核生物的基因调控
从酵母到哺乳动物转录调控的保守机制
活化子的ＤＮＡ结合与激活功能的分离
框１７－１　　双杂交实验
真核细胞调控蛋白使用一系列ＤＮＡ结合域，但ＤＮＡ识别的原理与细菌的相同
激活域结构不确定性
通过真核生物活化子募集结合基因的蛋白复合物
活化子募集转录机器到基因上
活化子也募集核小体修饰物来帮助转录机器结合到启动子上
在某些启动子上活化子募集另一个因子来进行有效的起始和延伸
远距作用：环与绝缘子
某些基因群的正确调控需要位点控制区域
框１７－２　　同一和不同染色体上远距离相互作用